

Epigenomic landscape of *Arabidopsis thaliana*

Minerva Susana Trejo-Arellano

*Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology
Uppsala*

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2019

Acta Universitatis agriculturae Sueciae

2019:9

Cover: "Wrong turn, and senescence ahead"

Drawing: María Guadalupe Trejo-Arellano

Layered colored pencil

Inspired by the *Epigenetic Landscape* by Conrad H. Waddington, 1957

ISSN 1652-6880

ISBN (print version) 978-91-7760-336-8

ISBN (electronic version) 978-91-7760-337-5

© 2019 Minerva S. Trejo-Arellano, Uppsala

Print: SLU Service/Repro, Uppsala 2019

Epigenomic landscape of *Arabidopsis thaliana*

Abstract

The basal unit of information storage in the cell is at the level of the DNA sequence. Dynamic packaging of DNA into chromatin allows additional regulatory layers. Two pairs of each histones H3, H4, H2A and H2B constitute an octamer around which about 146 base pairs of DNA are wrapped to build the basic unit of chromatin: the nucleosome. Furthermore, DNA nucleotides can be methylated and the amino-termini of histones tails can be covalently modified by a range of post-transcriptional modifications (PTMs), impacting on DNA accessibility and gene activity.

Compared to yeast and animals, the repertoire of histone PTMs in plants is less comprehensive. We identified two novel modifications on histone H3 that were not previously described in plants, but are evolutionarily conserved in high eukaryotes. Acetylation of lysine 36 (H3K36ac), which is enriched in euchromatin and seems to act as a binary indicator of transcription; and monomethylation of lysine 23 (H3K23me1) preferentially found in heterochromatin and along coding genes where it coexists with gene body CG DNA methylation. The spatiotemporal combination of histone modifications and DNA methylation comprises another layer of epigenetic control. Throughout development, changes in the epigenome are programmed to surpass developmental barriers. To address the configuration of the methylome during late stages of plant development, we profiled the DNA methylation landscape of darkness-induced senescent leaves of *Arabidopsis thaliana* and found modest methylation changes mainly localizing to transposable elements (TEs). Transcriptome changes revealed impairment of a module regulating chromatin conformation, suggesting that the observed changes in DNA methylation are a consequence of global chromatin rearrangements. Similarly, we observed DNA methylation changes localized to pericentromeric TEs in *Arabidopsis* plants lacking the Chromatin Assembly Factor 1 (CAF-1), a chromatin maintainer during replication and DNA repair. Therefore, disturbed chromatin packaging allows access of the DNA methylation machinery to regions of the genome otherwise inaccessible, highlighting the importance of chromatin structure to prevent faulty DNA modifications. Together, this thesis expanded our knowledge of the epigenetic landscape in *Arabidopsis* and integrated it into the regulatory networks that drive cell-fate decisions in plants.

Keywords: Epigenetics, chromatin, DNA methylation, histone modification, transcriptomics, ChIP-seq, Bisulphite sequencing

Author's address: Minerva S. Trejo-Arellano, SLU, Department of Plant Biology, Linnean Centre for Plant Biology, P.O. Box 7080, SE-750 07 Uppsala, Sweden

Configuración epigenómica de *Arabidopsis thaliana*

Resumen

La secuencia de ADN representa el primer nivel de almacenamiento de información en la célula. El empaquetamiento dinámico del ADN en la cromatina permite la existencia de niveles de regulación adicionales. Dos pares de cada una de las histonas H3, H4, H2A y H2B forman un complejo octamérico alrededor del cual se empaquetan 146 pares de bases de ADN para conformar la unidad básica de la cromatina: el nucleosoma. Además de que los nucleótidos del ADN pueden ser metilados, las colas amino-terminales de las histonas pueden ser blanco de modificaciones postranscripcionales (MPTs), impactando la accesibilidad del ADN y, en consecuencia, la expresión génica.

El repertorio de MPTs conocidas en histonas de plantas es aún limitado comparado con el de animales y levadura. En este trabajo identificamos dos modificaciones en la cola de la histona H3 que no habían sido descritas en plantas, pero están conservadas en eucariota. La acetilación en la lisina 36 (H3K36ac) está presente en la eucromatina donde actúa como marcador binario de la transcripción; y la monometilación de la lisina 23 (H3K23me1) se encuentra de manera preferente en la heterocromatina. La combinación espaciotemporal entre las MPT de histonas y la metilación del ADN constituyen otra capa de control epigenético. A lo largo del desarrollo, el epigenoma es modificado en el contexto de la diferenciación celular. En este trabajo se reporta el perfil de metilación de hojas senescentes de *Arabidopsis thaliana*. Los cambios de metilación en dicha condición avanzada del desarrollo son modestos y se encuentran localizados a lo largo de transposones. Descubrimos, además, la desregulación de un conjunto de genes involucrados en el mantenimiento de la cromatina, sugiriendo que las alteraciones observadas en el perfil de metilación son consecuencia de una reestructuración global de la cromatina. De manera similar, reportamos cambios localizados en el perfil de metilación de transposones y regiones pericentroméricas en el genoma de *Arabidopsis* carentes de una versión funcional de Chromatin Assembly Factor 1 (CAF-1), una chaperona de la cromatina durante la replicación y reparación del ADN. Estos resultados apuntan a que alteraciones en la cromatina permiten acceso a la maquinaria de metilación del ADN a regiones del genoma normalmente ocluidas, haciendo evidente la importancia de la correcta conformación de la cromatina. En conjunto, esta tesis contribuye a la descripción de la configuración epigenética del genoma de *Arabidopsis* y su impacto en las redes de regulación que controlan las transiciones de desarrollo en plantas.

Keywords: Epigenética, cromatina, metilación del ADN, modificaciones de histonas, transcriptoma, ChIP-seq, secuenciación bisulfito

Author's address: Minerva S. Trejo Arellano, SLU, Department of Plant Biology, Linnean Centre for Plant Biology, P.O. Box 7080, SE-750 07 Uppsala, Sweden

Dedication

To the moments and beloved ones
who flew away
while I was away, flying.

To my family.

To Lars.

Óyeme con los ojos, ya que están tan distantes los oídos.
Sor Juana Inés de la Cruz

Contents

List of publications	11
List of tables	17
List of figures	19
Abbreviations	21
1 Introduction	25
1.1 DNA methylation	27
1.1.1 Dynamics of DNA methylation	29
1.1.2 <i>De novo</i> DNA methylation in <i>Arabidopsis</i>	30
1.1.3 Maintenance of DNA methylation in <i>Arabidopsis</i>	33
1.1.4 DNA demethylation	33
1.1.5 Functional relevance of DNA methylation	34
1.2 Untangling the chromatin and its components	36
1.2.1 The nucleosome	36
1.2.2 Histone identity: variants and histone post-translational modifications	37
1.2.3 Histone methylation: writers, readers and erasers	41
1.2.4 Histone methylation and active chromatin	41
1.2.5 Histone methylation and repressed chromatin	42
1.2.6 Histone lysine demethylases	46
1.2.7. Histone acetylation: writers, readers and erasers	46
1.3 Chromatin assembly	50
1.4 Methods to study chromatin biology	52
1.4.1 Profiles of nucleosome occupancy and chromatin accessibility	53
1.4.2 Expanding the chromatin repertoire: Histone post-translational modifications and histone variants	54
1.4.3 DNA methylation landscapes	55
1.4.4 3D genome and intranuclear distribution	56
2 Aims of the study	59
3 Results and discussion	61

3.1 Expanding the histone language: H3K36ac and H3K23me1 are conserved histone modifications not previously described in plants	61
3.1.1 H3K36ac is enriched in euchromatin and colocalizes with activating histone marks	62
3.1.2 H3K23me1 dually marks eu- and heterochromatin but does not have a direct effect in gene expression	63
3.2 The <i>Arabidopsis</i> genome and chromatin configurations are robust to the lack of the replication dependent CAF-1	64
3.3 The <i>Arabidopsis</i> methylome is robust to disrupted chromatin	65
3.3.1 Transgenerational aggravation of CAF-1 mutants is accompanied by sequence-specific DNA methylation changes in <i>Arabidopsis</i>	65
3.3.2 The <i>Arabidopsis</i> methylome remains stable upon dark-induced senescence	66
4 Conclusions	69
5 Future perspectives	71
References	73
Popular science summary	91
Resumen de divulgación	93
Acknowledgements	95

List of publications

This thesis is based on the work contained in the following papers:

- I **Minerva S. Trejo-Arellano**[★], Saher Mehdi, Jennifer the Jonge Eva Dvorák Tomastíková, Claudia Köhler and Lars Hennig. Senescence causes localized changes in DNA methylation in *Arabidopsis* (manuscript)
- II Iva Mozgová[◇], Thomas Wildhaber[◇], **Minerva S. Trejo-Arellano**, Jiri Fajkus, Pawel Roszak, Claudia Köhler and Lars Hennig[★] (2018). Transgenerational phenotype aggravation in CAF-1 mutants reveals parent-of-origin specific epigenetic inheritance. *New Phytologist*, 220(3), 908-921.
- III **Trejo-Arellano MS**[◇], Mahrez W[◇], Nakamura M, Moreno-Romero J, Nanni P, Köhler C, Hennig L[★] (2017). H3K23me1 is an evolutionarily conserved histone modification associated with CG DNA methylation in *Arabidopsis*. *Plant Journal*, 90(2):293-303.
- IV Muñoz-Viana R, Wildhaber T, **Trejo-Arellano MS**, Mozgová I, Hennig L[★] (2017). *Arabidopsis* Chromatin Assembly Factor 1 is required for occupancy and position of a subset of nucleosomes. *Plant Journal*, 92(3), 363-374.
- V Walid Mahrez, **Minerva Susana Trejo Arellano**, Jordi Moreno-Romero, Miyuki Nakamura, Huan Shu, Paolo Nanni, Claudia Köhler, Wilhelm Grissem and Lars Hennig[★] (2016). H3K36ac Is an Evolutionary Conserved Plant Histone Modification That Marks Active Genes. *Plant Physiol*, 170(3), 1566-1577.

Paper II was first published in the New Phytologist Journal. The authors remain as copyright owners of the article. Hence, it is reproduced here.

Paper III is reproduced with the permission of the publisher under the licence number 4505360050467 granted to Minerva Susana Trejo Arellano, as author of the mentioned Wiley article.

Paper IV is reproduced with the permission of the publisher under the licence number 4505351475485 granted to Minerva Susana Trejo Arellano as author of the mentioned Wiley article.

Paper V was published under the “Copyright American Society of Plant Biologists” license. It is reproduced here under this agreement.

★ Corresponding author.

◇ Equal contribution

Additional publication produced during the course of the doctoral studies but not included in this thesis.

- I Mahrez W, Shin J, Muñoz-Viana R, Figueiredo DD, **Trejo-Arellano MS**, Exner V, Siretskiy A, Gruissem W, Köhler C and Hennig L[★] (2016). BRR2a Affects Flowering Time via FLC Splicing. *PLoS Genetics*, 12(4), 1-25.

★ Corresponding author.

The contribution of Minerva Susana Trejo-Arellano to the papers included in this thesis was as follows:

- I Planned, performed and interpreted the major part of the *in silico* experiments and wrote the largest part of the manuscript.
- II Suggested and performed the experiments to interpret the bisulphite sequencing data.
- III Planned, performed and interpreted the analysis of the epigenomic data sets. Wrote the final version of the manuscript with input from the other authors.
- IV Suggested and performed the experiments to interpret the genomic resequencing data.
- V Planned, performed and interpreted the data from H3K36ac ChIP-seq. Took major responsibility in the writing of the manuscript.

List of tables

Table 1. Histone modifications described so far in <i>Arabidopsis</i> and their corresponding histone writers and erasers	48
---	----

List of figures

Figure 1. Cytosine methylation and demethylation.	29
Figure 2. Canonical and non-canonical <i>de novo</i> RNA-directed DNA methylation (RdDM) pathway.	32
Figure 3. H3K36ac enrichment with respect to the TSS of genes.	62
Figure 4. H3K23me1 enrichment along TEs and genes.	63
Figure 5. Transgenerational aggravation of <i>Arabidopsis</i> phenotype in <i>fas2</i> mutants.	64
Figure 6. Context specific changes in the DNA methylation landscape of <i>fas2</i> mutants.	66
Figure 7. Global downregulation of genes encoding for members maintaining chromatin conformation and DNA methylation landscape.	67
Figure 8. The methylome of dark-induced senescent leaves remained largely stable in the three methylation contexts.	68

Abbreviations

5mC	5-methylcytosine
AdoMet	S-adenosyl-L-methionine
AG	AGAMOUS
AGO	ARGONAUTE
AP	APETALA
ASHH	ASH1-HOMOLOGUE
ASHR	ASH1-RELATED
ATX	ARABIDOPSIS TRITHORAX-LIKE
ATXR	ARABIDOPSIS TRITHORAX-RELATED
bp	Base pair(s)
CAC	Chromatin Assembly Complex
CAF-1	Chromatin Assembly Factor 1
ChIP	Chromatin immunoprecipitation
CLSY1	SNF2 DOMAIN-CONTAINING PROTEIN CLASSY 1
CMT3	CHROMOMETHYLASTE 3
DCL3	DICER-LIKE PROTEIN 3
DHSs	DNase I hypersensitive sites
DME	TRANSCRIPTIONAL ACTIVATOR DEMETER
DML2	DEMETER-LIKE PROTEIN 2
DML3	DEMETER-LIKE PROTEIN 3
DMS3	DEFECTIVE IN MERISTEM SILENCING 3
DNase I	Deoxyribonuclease I
DNMT1	DNA (cytosine-5)-methyltransferase 1
DRD1	DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1
DRM2	DOMAINS REARRANGED METHYLASE 2
dsRNA	Double stranded RNA
E(Z)	Enhancer of Zeste
ESC	Extra Sex Comb

<i>FAS1</i>	<i>FASCIATA1</i>
<i>FAS2</i>	<i>FASCIATA2</i>
FLC	FLOWERING LOCUS C
FT	FLOWERING LOCUS T
gbM	Gene-body methylation
HAT	Histone acetyltransferases
HDAC	Histone deacetylases
HMT	Histone methyltransferases
JmjC	Jumonji C
KYP	KRYPTONITE
LDL	Lysine Specific Demethylases (LSD)- like
LHP1	LIKE HETEROCHROMATIN PROTEIN 1
MBD	Anti-methylcytosine binding proteins
MET1	METHYLTRANSFERASE 1
MNase	Micrococcal nuclease
MPT	Modificaciones postranscripcionales
MS	Mass spectrometry
MSI1	MULTICOPY SUPPRESSOR OF IRA1
NCP	Nucleosome core particle
NGS	Next generation sequencing
NRPE1	DNA-DIRECTED POL V SUBUNIT 1
nt	Nucleotide
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase chain reaction
PI	PISTILATA
Pol IV	RNA POLYMERASE IV
PRC2	POLYCOMB REPRESSIVE COMPLEX 2
PRMT	Protein arginine methyltransferases
PTGS	Post-transcriptional gene silencing
PTM	Post-translational modifications
RdDM	RNA-directed DNA methylation
RDM1	RNA-DIRECTED DNA METHYLATION 1
rDNA	Ribosomal DNA
RDRP2	RNA-DEPENDENT RNA POLYMERASE 2
RNase III	Ribonuclease III
ROS1	REPRESSOR OF SILENCING 1
SDG	SET domain-containing groups
SHH1	SAWADEE HOMEODOMAIN HOMOLOGUE 1
siRNA	Small interfering RNAs
SRA	SET and RING finger-associated domain

ssDNA	Single-stranded DNA
ssRNA	Single strand RNA
SU(Z)12	Suppressor of Zeste 12
SUVH	SUPPRESSOR OF VARIEGATION 3-9 HOMOLOG
SUVR	SUPPRESSOR OF VARIEGATION 3-9 RELATED
TADs	Topologically associating domains
tasiRNA	Trans-acting siRNA genes
TSS	Transcriptional start site
VIM	VARIANT IN METHYLATION

1 Introduction

The current usage of the term *epigenetics* aims to gather our understanding of the mechanisms and modes of regulation that make it possible for cells carrying the same genetic identity to differentiate and fulfil specialized tasks within a living organism. This has led to the working definition of epigenetics as “*mitotically and/or meiotically heritable changes in gene function (and phenotypic outcomes) that cannot be explained by changes in the DNA sequence*” (Fincham, 1997). Similar to what happened with classical genetics, key observations made on plants have helped to build our understanding of epigenetic phenomena occurring without apparent changes to the DNA sequence. Barbara McClintock’s discovery of transposable elements (TEs) and description of their unstable activity in the 1940s and 1950s (McClintock, 1956), paved the way to build up some of the most detailed descriptions of epigenetic mechanisms controlling gene regulation [Reviewed in (Jones, 2005)]. Around the same time, Brink described paramutation in maize as the heritable influence in gene expression of one paramutable allele upon co-occurrence with its paramutagenic counterpart (Brink, 1956). The origin of epigenetic studies in mammals is inherently linked to the discovery of somatic inheritance of the “on” and “off” expression state of the X-chromosome in mouse (Lyon, 1961; Ohno *et al.*, 1959). This was followed by observations of uneven expression of sex and autosomal genes depending on their parental origin. It was work on insects that led to the term “chromosome imprinting” to describe the phenomenon in which parent-of-origin epigenetic marks along chromosomes were used by either of the parents to control genetic expression in the next generation (Cooper *et al.*, 1971; Crouse, 1960). Twenty years after the discovery and description of the first functional effects of epigenetic mechanisms, DNA methylation was

independently proposed as the signal responsible for those heritable functional changes (Sutter & Doerfler, 1980; Vardimon *et al.*, 1980; Holliday & Pugh, 1975; Riggs, 1975). Ever since the initial observations acknowledging DNA methylation as a key regulator of gene expression, a great deal of research has aimed to understand the establishment, maintenance and transmission of endogenous patterns of DNA methylation. Nevertheless, understanding the mechanisms in charge of removing methylated residues, which happens early in development and in the germline, is equally important. This is because the efficiency of DNA demethylation determines the transgenerational inheritance of the methylated cytosines that escape from demethylation events.

Before DNA methylation was explored as an epigenetic player, scaffold proteins were proposed to explain phenotypic differences observed among cells with the same karyotype (Stedman & Stedman, 1950). It was believed that histones defining a given phenotype were different from the histones in cells with distinct identity. Still, the discovery of histone variants was far away at that time. The first observations of a direct correlation between histone acetylation and gene activation (Allfrey & Mirsky, 1964), along with the discovery of other histone modifications, corroborated the existence of subtle differences that modulated the identity of core histones. The study of the nucleoprotein nature of the chromatin was exponentially facilitated with the resolution of the nucleosome, first at 7Å and later at 2.8Å, revealing the histone tails as key components of the histone octamer (Luger *et al.*, 1997). From then on, biochemical and genetic studies focused on complementing and characterizing the repertoire of histone modifications, histone variants and their functional impact. Equally important was the discovery of active mechanisms remodelling the accessibility of the chromatin and modulating gene expression (Peterson & Herskowitz, 1992; Yoshinaga *et al.*, 1992). The observation that the heterochromatin protein HP1 could bind the histone modification H3K9me2 and promote DNA methylation to spread silenced chromatin was the first hint towards the mechanistic and functional connection of nucleosome identity and DNA methylation (Bannister *et al.*, 2001). In the context of development, Polycomb group (PcG) protein-mediated H3K27me3 deposition is one of the most intensively studied mechanisms controlling the developmental program in animals and plants. Despite great technical advances achieved since the resolution of the nucleosome crystal

structure, the discovery and functional characterization of new histone modifications continues to be a challenging area of chromatin research.

A biological revolution started with the implementation of high-throughput or next generation sequencing (NGS) technologies that boosted our understanding of the molecule of life. It is evident that beyond the mere nucleotide sequence, a range of additional chemical modifications, together with scaffolding and regulatory proteins comprising the chromatin, are necessary to modulate the potential stored in the DNA. Genome-wide epigenome profiling studies pinpoint the specific location of those chromatin marks, including, but not restricted to DNA methylation, non-coding RNA molecules, histone post-translational modifications and histone variants. Nevertheless, individual epigenomic experiments reflect only the static configuration of the chromatin at the moment of the snapshot. Cross-correlation of those datasets allows the reconstruction of a dynamic chromatin regulation, providing a comprehensive profile of chromatin changes during development. Exploiting the potential offered by these technical advances, the current thesis describes our efforts to fill current knowledge gaps in the interplay of epigenetic mechanisms controlling plant development and its response to environmental changes, using the crucifer flowering plant *Arabidopsis thaliana* as a model.

1.1 DNA methylation

The addition of a methyl moiety to the fifth carbon of cytosine is known as 5-methylcytosine (5mC, Figure 1) or DNA methylation. DNA methylation was the first epigenetic signal to be recognised as such. A failed DNA-restriction experiment by HpaII (5'-CCGG-3') of the adenovirus type 12 (Ad12) genome integrated in hamster cells hinted to the existence of factors other than the DNA sequence affecting the potential for restriction digest by hindering the recognition of the restriction site (Sutter & Doerfler, 1980; Sutter *et al.*, 1978). Further experiments showed that the Ad12 genome became highly *de novo*-methylated upon integration into the host genome. This was the first observation documenting the specific targeting of foreign DNA by the host *de novo* DNA methylation machinery [Reviewed in (Doerfler, 2008)]. Notwithstanding, the real connection between DNA methylation and epigenetic phenomena was established

by showing an inverse correlation between promoter activity and its methylation status (Sutter & Doerfler, 1980). Contemporary studies to those in Ad12 addressed the role of DNA methylation in transgenerational inheritance of epigenetic states. The model of choice was X-chromosome inactivation; the random selection of one X chromosome to be silenced and the inheritance of the silenced status to daughter cells. Because there were no apparent changes in the DNA sequence, DNA methylation was proposed as the main epigenetic mark accounting for X chromosome inactivation (Holliday & Pugh, 1975; Riggs, 1975). In order for DNA methylation to be heritable, it is necessary that methylation prone sequences are palindromic. This would make it possible for DNA methylation enzymes to quickly copy the methylation status from the parental strand to the newly synthesized complementary strand at the palindrome site and thereby faithfully transmitting the methylation status onto the next generation (Bird, 2002).

DNA methylation is a widely spread mechanism to epigenetically control gene expression and genome integrity. Intuitively, steric hindrance by DNA methylation impairs the binding of the transcriptional machinery to promoters, enhancers and other regulatory DNA motifs. Furthermore, DNA-methylation mediated heterochromatin formation protects the integrity of the genome by preventing the uncontrolled jumping of TEs. In this regulatory context, DNA methylation can also interact with histone modifications to reinforce transcriptional silencing (He *et al.*, 2011). The bulk of DNA methylation contributing to heterochromatinization resides outside of gene bodies (Zhang *et al.*, 2006). Recent evidence in animals and plants, however, suggests that DNA methylation exists in gene bodies of long protein-coding genes, presumably, to stabilize the processivity of the DNA polymerase, enhancing gene expression (Muyle & Gaut, 2018; Yang *et al.*, 2014; Lister *et al.*, 2008; Zhang *et al.*, 2006). Nevertheless, some plant species can be completely devoid of gene body DNA methylation without showing an apparent effect (Bewick & Schmitz, 2017; Niederhuth *et al.*, 2016). Ultimately, DNA methylation and its transcriptional and structural outcomes reflect the integrated homeostasis of *de novo* methylation, maintenance of DNA methylation and active demethylation.

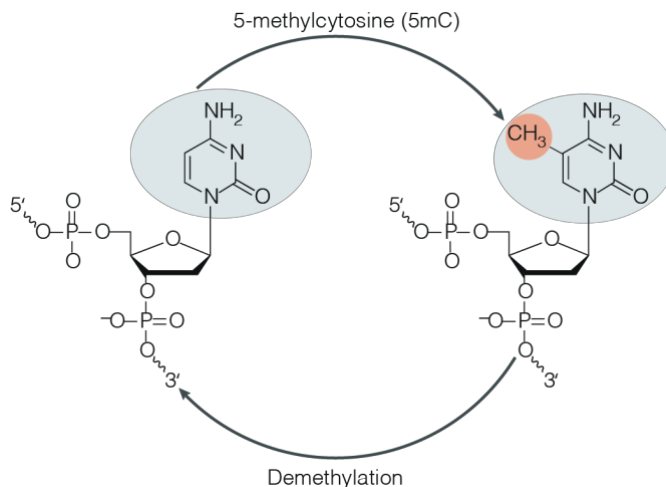


Figure 1. Cytosine methylation and demethylation. Scheme adapted and modified from (Zhang *et al.*, 2018)

1.1.1 Dynamics of DNA methylation

DNA methylation occurs in organisms from bacteria to eukaryotes (Blow *et al.*, 2016; Zemach & Zilberman, 2010; Noyer-Weidner & Trautner, 1993). In mammals, most of the DNA methylation is found in the symmetric CG context (Lister *et al.*, 2009), whereas plant genomes have the potential to be methylated in three different cytosine contexts: symmetric CG and CHG, as well as asymmetric CHH, where H can be A, T or G (Lister *et al.*, 2008). Conserved DNA methyltransferases catalyse the nucleophilic attack by which the cytosine 5th-carbon (5C) attacks the methyl group from the substrate S-adenosyl-L-methionine (AdoMet). Subsequently, the 5C proton is eliminated via β -elimination (Yang *et al.*, 2013; Vilkaitis *et al.*, 2001). In plants, and to a lower extent also in mammals, the RNA-directed DNA methylation (RdDM) pathway is crucial for *de novo* DNA methylation (Law & Jacobsen, 2010). Demethylation can be either passive, by repression of the methylation machineries and consequent genomic reduction of DNA methylation levels, or active (Figure 1). Active demethylation in plants is mediated by 5mC DNA glycosylases that specifically excise the methylated cytosine (Hsieh *et al.*, 2009; Zhu, 2009; Penterman *et al.*, 2007). This is in contrast to mammalian DNA demethylation, which requires the active deamination of the

5mC and further excision of the 5-hydroxymethylcytosine by the base-excision repair mechanism (He *et al.*, 2011). Despite the central role of DNA methylation in epigenetic silencing mechanisms, absence of functional components of the machineries regulating DNA methylation is not necessarily lethal in all plant species. The importance of DNA methylation in plants increases with the complexity of the genome, which is linearly correlated with the abundance of repetitive sequences whose mobilization threatens genome integrity (Bräutigam & Cronk, 2018). For instance, *Arabidopsis met1* mutants are viable, while *met1* in rice is seedling lethal (Hu *et al.*, 2014).

1.1.2 *De novo* DNA methylation in *Arabidopsis*

RNA-directed DNA methylation (RdDM) mediates *de novo* methylation in plants in all three sequence contexts (Cao & Jacobsen, 2002b). The key players of this pathway are RNA molecules in the form of small interfering RNAs (siRNA). siRNAs act at multiple levels of the pathway: either as triggers or as sequence-specific guides of the DNA methylation machinery to scaffold RNAs that complement the target DNA loci to be methylated (Matzke *et al.*, 2015; Jones *et al.*, 1999). The canonical RdDM pathway is triggered by the production of 24-nucleotide (nt) small RNAs by the plant specific RNA POLYMERASE IV (Pol IV) (Figure 1) (Herr *et al.*, 2005). RNA-DEPENDENT RNA POLYMERASE 2 (RDRP2) uses Pol IV transcripts as scaffolds to generate double stranded RNA (dsRNA), which is cleaved into siRNAs by DICER-LIKE PROTEIN 3 (DCL3) (Herr *et al.*, 2005; Xie *et al.*, 2004). The DCL-dependent siRNAs are then loaded into ARGONAUTE (AGO) proteins, mainly AGO4 and AGO6, and guide AGOs to the complementary nascent Pol V transcript (Duan *et al.*, 2015; Eun *et al.*, 2011; Zilberman *et al.*, 2003). In order to be used as a scaffold, the nascent Pol V-transcript is retained in the chromatin by RRP6-LIKE 1 (RRP6L1) (Zhang *et al.*, 2014). The siRNA::scaffold RNA pairing is stabilized by the INVOLVED IN DE NOVO 2 (IDN2)-IDN2 PARALOGUE (IDP) complex (Ausin *et al.*, 2009). siRNA-loaded AGO4/6 physically interacts with the DNA methyltransferase DOMAINS REARRANGED METHYLASE 2 (DRM2) and recruits it to catalyse *de novo* DNA methylation in the three cytosine contexts (Zhong *et al.*, 2014). RNA-DIRECTED DNA METHYLATION 1 (RDM1) is able

to associate with AGO4 and DRM2 and binds single-stranded DNA (ssDNA), assisting on the DRM2-mediated DNA methylation (Zhong *et al.*, 2014). In addition to the sequence-specific pairing of the siRNA and the scaffold transcript, the RdDM recognition and methylation protein complex is further stabilized by protein interactions between AGO4 and the AGO hook-containing proteins DNA-DIRECTED Pol V SUBUNIT 1 (NRPE1, the largest Pol V subunit) and RDM3, a Pol V-associated putative transcription elongation factor (Zhang *et al.*, 2014; Bies-Etheve *et al.*, 2009; He *et al.*, 2009).

The intrinsic configuration of the chromatin can prime the recruitment of Pol IV and Pol V, reinforcing the epigenetic state of the chromatin at RdDM target loci. SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1) recruits Pol IV and binds to H3K9me2 through its Tudor domain (Law *et al.*, 2013). SHH1 also interacts with the chromatin remodelling complex SNF2 DOMAIN-CONTAINING PROTEIN CLASSY I (CLSY1), which in association with Pol IV is required for the production of Pol IV-siRNAs (Zhang *et al.*, 2018). The chromatin remodelling complex DDR makes the chromatin accessible to be transcribed by Pol V that produces the scaffold RNAs (Kanno *et al.*, 2004). DDR is formed by the chromatin remodelling protein DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), a chromosome structure maintenance protein and RDM1 (Greenberg *et al.*, 2011; Law *et al.*, 2010). In addition to the recognition of H3K9me2-poised chromatin, pre-existing DNA methylation can also recruit the RdDM machinery. This is possible through the physical interaction of DDR with the proteins SUVH2 and SUVH9; SU(VAR) 3-9 histone methyltransferases impaired in their catalytic activity (Johnson *et al.*, 2014). SUVH2 and SUVH9 are required of proper Pol V occupancy and since their SET and RING finger-associated domain (SRA) recognizes methylated cytosines, they were proposed to recruit Pol V to already methylated DNA (Liu *et al.*, 2014). Regardless, SUVH9 can be independently recruited to unmethylated DNA by DNA binding proteins and establish DNA methylation and gene silencing (Johnson *et al.*, 2014).

In addition to the canonical RdDM pathway, in the non-canonical pathway, polyadenylated ssRNA Pol II transcripts are converted into dsRNA by RDR6 (Figure 2) [Reviewed in (Cuerda-Gil & Slotkin, 2016)]. DCL2 and DCL4 process the dsRNA into 21-22 nt siRNAs, which are loaded into AGO6 and guide the DRM2-RdDM machinery

to methylate the target loci (McCue *et al.*, 2015; Mari-Ordonez *et al.*, 2013; Nuthikattu *et al.*, 2013; Wu *et al.*, 2012).

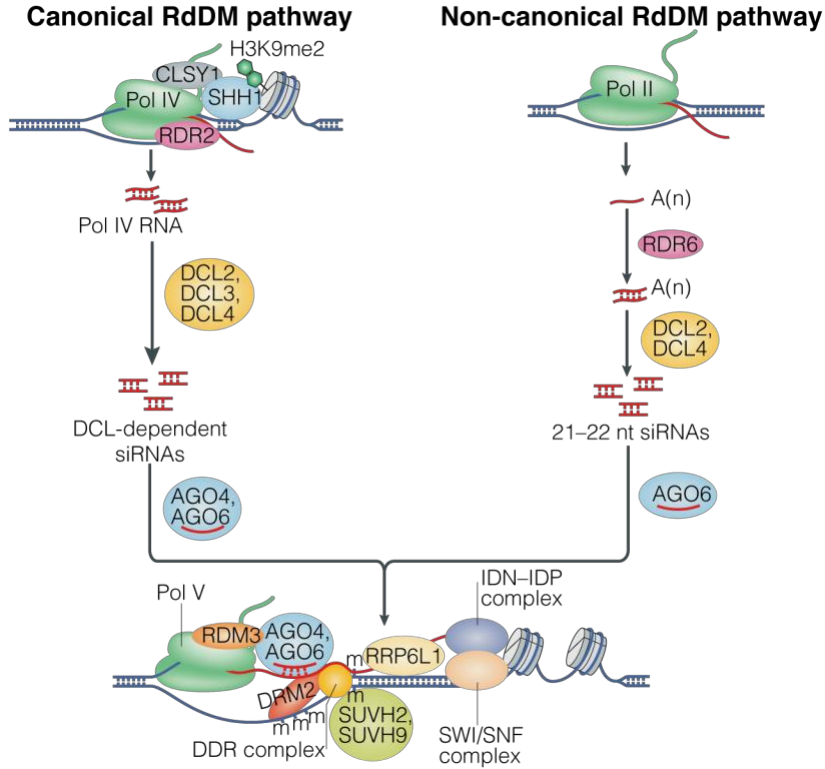


Figure 2. Canonical and non-canonical *de novo* RNA-directed DNA methylation (RdDM) pathway. Scheme adapted and modified from (Zhang *et al.*, 2018).

The majority of the small RNA molecules detected by genome-wide sequencing experiments are 24 nt in length (Tabara *et al.*, 2018). Nevertheless, in the *dcl* quadruple mutant that is almost fully depleted in 24-nt siRNAs, many RdDM target loci retain their DNA methylation status, indicating the existence of DCL-independent RdDM (Yang *et al.*, 2016), possibly assisted by the nuclease activity of endogenous ribonuclease III (RNase III) enzymes other than DCLs.

1.1.3 Maintenance of DNA methylation in *Arabidopsis*

Context-dependent methyltransferases ensure the transmission of the DNA methylation status to the daughter DNA strand after replication [Reviewed in (Law & Jacobsen, 2010)]. CG methylation is maintained by METHYLTRANSFERASE 1 (MET1), an ortholog of the mammalian DNA (cytosine-5)-methyltransferase 1 (DNMT1) (Liu *et al.*, 2012; To *et al.*, 2011). MET1 is recruited to the DNA by VARIANT IN METHYLATION (VIM) proteins, where it recognises hemi-methylated CG contexts and methylates the unmodified cytosine of the newly synthesized strand (Woo *et al.*, 2007). CHROMOMETHYLASE 3 (CMT3) is the main methyltransferase in charge of maintaining CHG methylation (Cao & Jacobsen, 2002a). The chromodomain and the bromo-adjacent homology domain (BAH) of CMT3 bind to H3K9me2 (Enke *et al.*, 2011; Johnson *et al.*, 2008). The presence and recognition of H3K9me2 is required for the proper CHG methylation maintenance by CMT3, creating a positive feedback loop in which H3K9me2 and CHG methylation reinforce each other (Du *et al.*, 2014). CMT2 also maintains CHG methylation but to a much lesser extent than CMT3 (Stroud *et al.*, 2014). The RdDM pathway preferentially targets evolutionarily young TEs, repetitive sequences interspersed in euchromatin, and the edges of long TEs (Huettel *et al.*, 2006). On the other hand, H1-containing heterochromatin inhibits the accessibility to RdDM. CHH methylation in this chromatin context is catalysed by CMT2 with the help of the chromatin remodeller DDM1 (Stroud *et al.*, 2014; Zemach *et al.*, 2013).

1.1.4 DNA demethylation

Once established, DNA methylation patterns are dynamically maintained throughout plant development [Reviewed in (Law & Jacobsen, 2010)]. During early developmental transitions of *Arabidopsis*, genomic methylation is partly erased and reprogrammed in the vegetative nucleus of pollen and in the central cell of the female gametophyte (Calarco *et al.*, 2012; Slotkin *et al.*, 2009). This is likely required to transmit epigenetic information and thus ensure genomic integrity (Ibarra *et al.*, 2012; Slotkin *et al.*, 2009). DNA demethylation is also crucial to activate the expression of imprinted genes (Li *et al.*, 2018). The *Arabidopsis* genome encodes four 5mC glycosylases: DEMETER-LIKE PROTEIN 2 (DML2), REPRESSOR OF

SILENCING (ROS1), TRANSCRIPTIONAL ACTIVATOR DEMETER (DME), and DEMETER-LIKE PROTEIN 2 (DML2/3) (Li *et al.*, 2018). *DME* is expressed in the central cell of the female gametophyte and required for the activation of maternal alleles of imprinted genes (Huh *et al.*, 2008). *DME* is also active in the vegetative cell of pollen, which may generate mobile sRNAs that enforce silencing of TEs in sperm (Calarco *et al.*, 2012). *ROS1*, *DML2* and *DML3* are expressed in somatic tissues. *DME*, *DML2* and *DML3* demethylate all three sequence contexts, particularly at loci with improper methylation (Ortega-Galisteo *et al.*, 2008). *ROS1* mainly counteracts methylation deposited by the RdDM pathway, but it can also demethylate RdDM-independent loci (Lei *et al.*, 2015; Williams *et al.*, 2015). These loci are hypermethylated in *ros1* but not in the double mutant *ros1 nrpd1*, revealing that *ROS1* and the RdDM pathway target the same loci (Tang *et al.*, 2016). Moreover, expression of *ROS1* is positively regulated by the RdDM pathway, suggesting that *ROS1* functions as an epigenetic rheostat, adjusting the level of demethylase activity to alterations of methylation (Williams *et al.*, 2015). The feedback between *ROS1* and RdDM is coordinated by a DNA methylation monitoring sequences (MEMS) that regulates *ROS1* expression (Lei *et al.*, 2015). MEMS is a 39 bp region in the promoter of *ROS1* targeted by MET1 and RdDM. Methylation of MEMS is required for *ROS1* expression, and mutants with a non-functional *ROS1* have a hypermethylated MEMS that display *ROS1* overexpression. In the *ROS1* promoter, just upstream of MEMS, there is a helitron transposon which may serve as a trigger to spread DNA methylation in the *ROS1* promoter (Lei *et al.*, 2015). Disruption of RdDM-mediated *ROS1* expression leads to widespread depletion of DNA methylation, especially in euchromatin, and abnormal phenotypes that progressively worsen across generations (Williams & Gehring, 2017).

1.1.5 Functional relevance of DNA methylation

Genome-wide studies have paved the way to understand the role of DNA methylation in the expression of the genome (Stroud *et al.*, 2013). In plants, heterochromatic regions are heavily methylated in all cytosine contexts. *Arabidopsis* and tomato pericentromeric heterochromatin, however, display preference to methylate CAA, CTA and CAT contexts, and low methylation levels at CCC (Gouil &

Baulcombe, 2016). Mutants of the DNA methylation machinery lead to TE reactivation. Double mutants for *MET1* and *CMT3* display global reduction of CG and CHG methylation and concomitant TE reactivation. This increase in TE expression is not observed in either of the single mutants (Mirouze *et al.*, 2009; Kato *et al.*, 2003). The *nrrpd1* mutant shows heat-stress sensitive remobilization of the retrotransposon *ONSEN*. However, mutants in *CMT2* or components of the RdDM pathway do not show the same effect (Ito *et al.*, 2011), suggesting the existence of additional factors other than just CHH methylation repressing uncontrolled transposition.

The RdDM pathway is not only required to repress TE activity, but it is also crucial to prevent chromosomal interactions, as made evident by increased interactions in mutants of the RdDM pathway. The model that has been proposed is that RdDM prevents long-range interactions between regulatory sites and genes that are both targeted by the RdDM pathway (Rowley *et al.*, 2017).

Gene-associated DNA methylation is generally found at the promoter regions or within the transcribed gene body (gbM); however, the regulatory potential of methylation differs depending on whether it is localized in the promoter or in the gene body. With some exceptions, such as the case of *ROS1* (Lei *et al.*, 2015), methylation of the promoter will prevent the binding of transcriptional regulators, rendering the gene as silenced. In many instances, methylation in promoters is a consequence of methylation primarily aimed to silence TEs in the neighbourhood of functional genes. How methylated promoters can overcome the repressive effect of methylation and allow transcription initiation remains to be shown.

About one third of *Arabidopsis* genes carry gbM, which is preferentially deposited in CG contexts and dependent on *MET1* and *CMT3*. A great debate exists around the relevance of gbM, both in plants and animals (Bewick & Schmitz, 2017; Zilberman, 2017; Bewick *et al.*, 2016). Since gbM is generally found towards the 3'-end of long genes (Zhang *et al.*, 2006), it has been proposed to prevent aberrant transcription initiation at internal cryptic promoters (Neri *et al.*, 2017). In another scenario, the presence of the H2A.Z variant in the gene body correlates with gene responsiveness and gbM seems to be mutually exclusive with H2A.Z deposition. As such, gbM has been proposed as a strategy to reduce expression variability by excluding H2A.Z (Zilberman *et al.*, 2008). While gbM is present in most angiosperms, exceptions have been identified that lack gbM (Muyle &

Gaut, 2018). The emergence of gbM has been correlated with the evolution of CMT3 (Bewick *et al.*, 2016); however, the mechanism by which CMT3 catalyses CG methylation remains to be established. Furthermore, natural gbM variation has been found in *Arabidopsis* accessions, but it does not correlate with variation in mRNA levels, suggesting that gbM is a direct or indirect consequence of transcription rather than a cause (Kawakatsu *et al.*, 2016). Therefore, the functional role of gbM remains elusive.

1.2 Untangling the chromatin and its components

The basal unit of information in the cell is at the level of the DNA sequence. The dynamic packaging of the DNA into a chromatin structure allows multiple levels of regulation. Historically, chromatin has long been regarded as an efficient evolutionary invention to pack a huge amount of genomic information into the reduced nuclear compartment of a cell. This pragmatic approach to chromatin biology has been boosted by the gradual discovery and characterization of the building blocks of chromatin. The biochemical characterization of the nucleoprotein nature of chromatin, coupled to the power offered by next generation sequencing, has made it possible to profile the multiple levels of chromatin composition and its concomitant influence on the transcriptional readout. It is clear now that beyond its role as a compacting strategy, the nucleoprotein configuration of the chromatin is crucial to dynamically regulate the expression of the genome.

1.2.1 The nucleosome

The lowest level of chromatin organization consists of the wrapping of 1.7 turns, equivalent to approximately 146 base pairs (bp), of superhelix DNA around the histone octamer made of four heterodimers: two of H3-H4 and two of H2A-H2B (Luger *et al.*, 1997). This nucleoprotein complex of a histone octamer serving as a scaffold for the 146 bp of DNA is known as the nucleosome core particle (NCP). It constitutes the primary building block of the chromatin and as such, it is found as the repeating packaging motif along the entire DNA molecule. In between the NCP, a variable length of linker DNA is associated to linker histone H1, making the DNA coiled around the octamer to go from 146 to 166 bp and completing

two full turns of DNA around the octamer (Tan & Davey, 2011). Within chromatin, the structure formed by the core particle together with histone H1 and linker DNA, is termed as the chromatosome and is traditionally considered as the fundamental unit of chromatin (Widom, 1998). The chromatosome was first revealed by digesting the chromatin with micrococcal nuclease. Prolonged digestion of the complex releases the 146 bp of DNA coiled around the NCP (Simpson, 1978).

The nucleosome, being the building block of chromatin, determines DNA accessibility. Meaning that it has to be modular enough to fulfil the packaging constraints required by the chromatin but at the same time ensure the timely binding of DNA sequence-specific binding proteins and RNA/DNA polymerases. This can be achieved by multiple strategies, including the turnover of canonical histones with functionally specialized histone variants; the homeostasis of writing and removal of post-translational modifications (PTM) to the histone tails and/or globular domains; translational displacement of the nucleosome, or even complete removal of the histone octamer. The energetic stability of the nucleosome and its potential to be slid or removed from the DNA is mainly determined by the identity of the histones composing the octamer, along with the concomitant PTM of histones and their interaction with the DNA sequence (Henikoff, 2008). Ultimately, intrinsic and extrinsic triggers of transcriptional response convey on the concerted action of histone readers and writers, and ATP-dependent chromatin remodelers, to regulate the exposure or occlusion of the DNA. This dynamic interplay leads to the formation of different chromatin territories of varying degree of condensation, with euchromatin as the relaxed transcription-prone state and heterochromatin as the compacted and silent one. The five chromosomes of *Arabidopsis* for instance, have a rather simplistic structure, with pericentromeric heterochromatin and euchromatic chromosome arms.

1.2.2 Histone identity: variants and histone post-translational modifications

The inherent need to store the DNA molecule into a constrained physical space is intrinsic to every form of living organism. To tackle this issue, evolution has developed architectural proteins that compress the DNA but at the same time allow access to functional

protein complexes. The delicate equilibrium between packaging and functional constraints explains the fact that only two classes of scaffolding proteins are found throughout the tree of life: HU proteins packaging the bacterial DNA and histones, packaging eukaryotic DNA (DNA from archaea can be packed by either of them) (Talbert & Henikoff, 2010).

Compared to the complex structure of the eukaryotic NCP, the archaeal nucleosome unit is much simpler. It is composed of proteins that contain the histone-fold domain but lacking tails. They assemble into a tetramer around which one single turn of right-handed superhelix DNA is wrapped (Reeve *et al.*, 1997). Nevertheless, structural features pose it as the ancestral particle of eukaryotic nucleosomes. For instance, the crystal structure of the archaeal tetramer superimposes the structure of the eukaryotic (H3-H4)₂, suggesting that eukaryotic nucleosomes evolved from an archaeal ancestor by duplicating the number of heterodimers to accommodate a second turn of left-handed superhelix DNA (Henneman *et al.*, 2018). Moreover, even though most of the archaeal histones are variants interchangeably loaded into the tetramer, some of them can be fused into dimers that occupy a specific position in the nucleosome. Similarly, the eukaryotic histones come together in dimers: H2A with H2B and H3 with H4 and superimpose with archaeal nucleosomes (Ammar *et al.*, 2012). Interestingly, H2A and H3 have widely diversified into histone variants. Whereas H2B and H4 display almost no diversification (Pusarla & Bhargava, 2005). This duality has allowed for the functional diversification of nucleosomes containing histone variants while still preserving the structural prerequisites of an architectural particle.

While the canonical histones are loaded during DNA replication, histone variants are incorporated during interphase [Reviewed in (Serra-Cardona & Zhang, 2018)]. H2A has two major variants: H2AX and H2A.Z. H2AX is loaded at sites that have undergone DNA damage and activated DNA repair pathways (van Attikum & Gasser, 2009). H2A.Z has been implicated in the regulation of the expression and maintenance of the genome and delimiting heterochromatin boundaries. Genome-wide profiling locates H2A.Z throughout the whole genome but preferentially immediately upstream of the transcriptional start site (TSS), presumably as a strategy to prevent methylation of the promoter and consequently, gene silencing (Zilberman *et al.*, 2008).

H3 has two variants in eukaryotes: CenH3, and H3.3. CenH3 is incorporated in centromeres, independently of the centromeric sequence. It is essential for the correct formation of kinetochores, the sites of attachment of the spindle microtubules during mitosis and meiosis, and thus, essential for correct chromosome segregation (Amor *et al.*, 2004). The canonical histone H3 is enriched in heterochromatin and other silent regions of the genome. Nucleosomes containing the canonical H3 are also found to carry silencing histone modifications such as H3K9me2 and H3K27me3 (Stroud *et al.*, 2012). H3.3, which differs from the canonical H3 in only three to four amino acids, is the substrate for the replication-independent nucleosome assembly pathway. A specialized group of histone chaperones, including HirA and ATRX/Daxx assist the incorporation H3.3 (Duc *et al.*, 2017; Elsaesser & Allis, 2010; Lewis *et al.*, 2010). In *Arabidopsis*, H3.3 containing nucleosomes are found at the promoters of active genes (Shu *et al.*, 2014).

Within the NCP, histones have two distinct domains: the histone-fold and the flexible histone tail protruding out of the nucleosome structure. The histone-fold domain is the most conserved and the most important for histone dimerization (Arents & Moudrianakis, 1995), while the tail is rather variable. All of the four core histones have an N-terminal tail, but only H2A possesses an additional globular C-terminal tail. The N-terminal tails are randomly coiled, allowing them to fit through the gyres of DNA. The high enrichment for basic amino acids such as lysine and arginine confers histone tails a positive charge, reinforcing the association with the DNA through electrostatic interactions (Davey & Richmond, 2002).

One of the major differences between ancestral and eukaryotic histones that have allowed diversification and evolving complexity, is the presence of N-terminal histone tails subject to extensive covalent post-translational modifications (Strahl & Allis, 2000). Among the described PTMs are acetylation, methylation, phosphorylation and ADP-ribosylation, as well as SUMOylation and ubiquitination. Further adding to the complexity promoted by histone PTMs, a single amino acid can carry more than one functional group: lysine can be mono-, di- or tri- methylated; whereas arginine can be mono- or dimethylated in a symmetric and asymmetric manner (Kouzarides, 2007). Histone modifications are typically detected using specific antibodies. Coupling the detection method with high-throughput genomic technologies allows their mapping and co-localization to

specific regions along the genome. Genome-wide combinatorial relationships of histone PTMs, histone variants and other epigenetic marks, discretize the chromatin into specific *chromatin states* that define topographical motifs preserving the integrity of the genome or concerting specialized patterns of gene expression (Sequeira-Mendes *et al.*, 2014; Roudier *et al.*, 2011). It is noteworthy that most of the post-translational modifications do not directly alter chromatin structure. They rather act as tethers of effector proteins, or “readers”, that recognize single or concerted histone modifications to couple downstream transcriptional regulatory events (Tropberger & Schneider, 2010). Acetylation is one of the few histone modifications that directly alters the stability of the nucleosome. Being a negative functional group, the addition of an acetyl group loosens the association between histones and the DNA and is generally associated with active transcription. Histone methylation is the most abundant modification. It can occur on the alkaline amino acids arginine (Arg, R) and lysine (Lys, K). Histone methylation is rather stable. Unlike acetylation, the addition of one or multiple methyl groups does not alter the charge of the modified residue. Neither does it lead to steric clashes with amino acids or the DNA. Hence, the effect of histone methylation on transcription is a consequence of the recognition by specific histone readers. Therefore, histone methylation can be associated both with activation and repression of transcription, depending on the histone, the modified amino acid and the degree (mono-, di-, tri-) of methylation. For instance, in *Arabidopsis*, genome-wide analysis showed that H3K4me3 and H3K36me3 are enriched along actively transcribed genes (Zhang *et al.*, 2009; Xu *et al.*, 2008), H3K27me3 is associated with repressed genes, whereas H3K9me2 is generally acknowledged as a marker of constitutively silenced heterochromatin (Jackson *et al.*, 2004). Additionally, because of divergent histone readers, the same histone modification might have a different effect in plant or animal genomes (Loidl, 2004). Methylation of H3K4 and H3K36 are highly conserved as euchromatic marks among eukaryotes, whereas heterochromatin indexing by methylation at H3K9, H3K27 and H3K20 is more variable. Immunostaining of H3K4me2 among 24 different species invariably found it to be enriched in euchromatic regions. The distribution of H3K9me2 on the other hand, seemed to correlate with the size of the genome: in species with a genome size of at most 500 Mb the signal was predominantly detected in heterochromatic chromocenters, but in

larger genomes it was found dispersed all over the nuclei (Houben *et al.*, 2003). The prevalent silencing mark in mammals is H3K9me3. This contrasts with the prevalence of heterochromatic mono- and di-H3K9 and H3K27 methylation in *Arabidopsis*, since trimethylated lysine is more robust than mono- and di- methylated ones (Fuchs *et al.*, 2006). In plants, methylation has been found in K4, K9, K27, K36 and K79 of histone H3 and K20 of histone H4. Arginine methylation has been identified at positions R2, R17 and R26 of histone H3 and R3 of histone H4 [Reviewed in (Liu *et al.*, 2010a)]. Because of the scope of the present thesis, the next sections will focus on the role of histone modifications in plants.

1.2.3 Histone methylation: writers, readers and erasers

“Writers” of histone lysine methylation are known as histone methyl transferases (HMTs). They all have a conserved catalytic SET domain (Thorstensen *et al.*, 2011), except for H3K79 methylation, whose deposition is catalyzed by Dot1, a family of non-SET domain containing HMTs. However, the *Arabidopsis* genome seems to be depleted of H3K79 methylation and it does not encode for any Dot1-like homolog. The *Arabidopsis* genome encodes for a group of SET-domain containing proteins (Thorstensen *et al.*, 2011), collectively known as SET-domain-containing groups (SDGs). Because the first SDG proteins were identified in *Drosophila*, SDGs in *Arabidopsis* are further subclassified based on their homology to the different *Drosophila* SET domain proteins (Pontvianne *et al.*, 2010):

- 1) Enhancer of Zeste [E(Z)] homologues
- 2) ARABIDOPSIS TRITHORAX-LIKE (ATX) or TRITHORAX-RELATED (ATXR)
- 3) ASH1-HOMOLOGUE (ASHH) or ASH1-RELATED (ASHR)
- 4) SUPPRESSOR OF VARIATION 3-9 HOMOLOG (SUVH) 10 genes in *Arabidopsis* or RELATED (SUVR)

1.2.4 Histone methylation and active chromatin

In *Arabidopsis*, H3K4 and H3K36 methylation is associated with gene expression. The deposition of mono-, di- or tri- methyl moieties on both, the fourth lysine of H3 (H3K4) and the 36th lysine of H3 (H3K36) is tightly linked. Different SDGs from the ATX/ATXR and

ASHH/ASHR families have the potential to methylate both, H3K4 and H3K36. SDG25/ATXR7 can mono-, di- and trimethylate H3K4. SDG25 loss of function causes early flowering due to suppression of *FLOWERING LOCUS C (FLC)* (Berr *et al.*, 2009). SDG2/ATXR3 is the most active H3K4 methyltransferase. It can deposit di- and trimethylation in H3K4. Lack of SDG2 causes pleiotropic defects during sporophytic and gametophytic development (Berr *et al.*, 2010; Guo *et al.*, 2010). SDG27/ATX1 can also trimethylate H3K4. It is required for proper regulation of flowering time and identity of floral organs by regulating the activation of *FLC* (Pien *et al.*, 2008; Alvarez-Venegas *et al.*, 2003). SDG4/ASHR3 trimethylates H3K4 and H3K36 and is involved in the correct development of stamen and pollen (Thorstensen *et al.*, 2008). SDG30/ATX2 dimethylates H3K4 (H3K4me2) but loss of SDG30 does not cause any obvious phenotype.

From genome-wide studies we know that H3K4me1 does not seem to have a direct effect on transcription, but its presence colocalizes with CG gene body methylation towards the 3'-end of long and actively transcribed genes. H3K4me2 and H3K4me3 on the other hand are found along promoters and the 5'-end of highly expressed genes (Zhang *et al.*, 2009).

H3K36me3 is enriched in the 5'-end of genes and is associated with gene activity. H3K36me2 is rather found towards the 3'-end, where it assists correct transcriptional elongation. SDG25/ATXR7, SDG4/ASHR3, and SDG26/ASHH1, that mainly methylate H3K4, can also methylate H3K36. However, the major H3K36 methyltransferase is SDG8/ASHH2. Impairment of the catalytic activity of SDG8 causes defects in flowering and pollen tube formation (Xu *et al.*, 2008).

1.2.5 Histone methylation and repressed chromatin

H3K9 methylation is catalysed by SUPPRESSOR OF VARIEGATION 3-9 methyltransferases. In plants, H3K9 is predominantly mono- and dimethylated. H3K9me1 and H3K9me2 mark genomic regions that are constitutively silenced in heterochromatic domains, but both marks are also found in repetitive sequences interspersed with euchromatin. H3K9me3, in contrast, is found in euchromatin as an activating mark targeting TSSs (Charron *et al.*, 2009). SUVH4/KRYPTONITE (KYP) was first discovered in two independent studies as a mutant releasing silencing of endogenous

silenced loci (Jackson *et al.*, 2002; Malagnac *et al.*, 2002). Importantly, those initial screenings already described altered patterns of cytosine methylation upon disturbed H3K9me2, uncovering the link between histone methylation and DNA methylation. This was later supported by the loss of H3K9 methylation in mutants of the main CG methyltransferase *MET1* (Tariq *et al.*, 2003). H3K9 is also methylated by SUVH5 and SUVH6. The triple mutant *suvh4 suvh5 suvh6* is depleted of H3K9 mono- and dimethylation and has a concomitant reduction of non-CG methylation (mimicking the *cmt3* phenotype), thereby strengthening the link between H3K9me2 and non-CG methylation (Ebbs & Bender, 2006). The role of SUPPRESSOR OF VARIATION 3-9 homologs in *Arabidopsis* is less well defined. SUVR5 was found to interact with the nucleosome remodeling factor AtSWP1 and to promote the downregulation of *FLC* by deposition of H3K9me2 and H3K27me3 (Caro *et al.*, 2012). SUVR4 promotes full methylation of H3K9me1 to H3K9me2 in transposon-rich heterochromatin (Veiseth *et al.*, 2011).

H3K27 monomethylation (H3K27me1) is deposited by TRITHORAX-RELATED protein 5 (ATXR5) and ATXR6 in *Arabidopsis* in a replication dependent manner (Jacob *et al.*, 2010; Raynaud *et al.*, 2006). Double *atxr5 atxr6* mutants display reduced H3K27me1 levels, partial heterochromatin decondensation and reactivation of TEs and other repetitive sequences in the centromeres. Importantly, H3K9me2 levels do not seem to be affected in the double mutant, suggesting independent involvement of H3K9me2 and H3K27me1 in heterochromatin repression (Jacob *et al.*, 2009). Steric hindrance prevents ATXR5 and ATXR6 from methylating nucleosomes containing the replication-independent H3.3 histone variant, thereby protecting H3.3-enriched genes against heterochromatinization during DNA replication. ATXR5/6 are recruited to the replication fork during S phase through their interaction with PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA), restoring H3K27me1 at newly incorporated H3 nucleosomes and preventing over replication. These results highlight the potential of histone variants to poise specific chromatin states and directly modulate the enzymatic activity of histone writers (Jacob *et al.*, 2014).

H3K27me3, a well-studied modification in the context of the regulation of developmental transitions, is deposited by the POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) both in plants and in animals. In *Drosophila melanogaster*, PRC2 is composed by four

conserved subunits: p55, Suppressor of Zeste 12 [SU(Z)12], Extra Sex Comb (ESC) and Enhancer of Zeste [E(Z)]. The actual deposition of H3K27me3 is catalyzed by E(Z) and ESC.

The *Arabidopsis* genome encodes homologs for all the members of *Drosophila*'s PRC2. The five copies of MULTICOPY SUPPRESSOR OF IRA 1, MSI1–5, are p55 homologs. EMBRYONIC FLOWER 2 (EMF2), and VERNALIZATION 2 (VRN2) are the *Arabidopsis* homologs of Su(Z)12. FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) is an ESC homolog in *Arabidopsis* and three homologs of Enhancer of Zeste [E(Z)] are the main catalytic subunits of *Arabidopsis* PRC2 methyltransferase activity: CURLY LEAF/SDG1 (CLF), MEDEA/SDG5 (MEA) and SWINGER/SDG10 (SWN). *CLF* and *MEA* were the first histone lysine methyltransferase genes described in plants, emphasizing the role of chromatin modifications for proper regulation of developmental transitions (Jiang *et al.*, 2008; Wood *et al.*, 2006; Grossniklaus *et al.*, 1998; Goodrich *et al.*, 1997). *CLF* trimethylates H3K27 along the *FLC* gene body during vernalization (Schubert *et al.*, 2006; Wood *et al.*, 2006), as well as other developmentally important genes. Mutants in *CLF* display pleiotropic phenotypes including early flowering, aberrant floral organs and curled leaves, hence the name of the subunit (Jiang *et al.*, 2008; Kim *et al.*, 1998). In some instances, *SWN* acts redundantly with *CLF*, as revealed by severe enhancement of phenotypic abnormalities in *swn clf* double mutants and strongly depleted H3K27me3 (Zheng & Chen, 2011).

Three major PRC2 complexes regulate development in *Arabidopsis* via H3K27me3 deposition:

- 1) FIS-PRC2, composed of MEA, FIS2, FIE and MSI1, is required in the female gametophyte to suppress autonomous replication of the central cell and to regulate endosperm development after fertilization (Mozgova & Hennig, 2015; Guitton *et al.*, 2004; Kohler *et al.*, 2003a; Kohler *et al.*, 2003b; Luo *et al.*, 1999; Grossniklaus *et al.*, 1998; Chaudhury *et al.*, 1997; Ohad *et al.*, 1996)
- 2) EMF-PRC2, composed of EMF, CLF, FIE and MSI1, regulates sporophyte development and the transition from the vegetative to the reproductive life phase (Derkacheva *et al.*, 2013; De Lucia *et al.*, 2008; Jiang *et al.*, 2008; Schonrock *et al.*, 2006a;

Wood *et al.*, 2006; Chanvivattana *et al.*, 2004; Gendall *et al.*, 2001; Yoshida *et al.*, 2001).

- 3) VRN2-PRC2, composed of SWN/CLF, VRN2, FIE and MSI1, is partially redundant with EMF complex in regulating sporophyte development and the transition from vegetative to reproductive life. Furthermore, the VRN2-PRC2 complex mediates *FLC* silencing via H3K27me3 spreading during vernalization (De Lucia *et al.*, 2008; Greb *et al.*, 2007; Sung *et al.*, 2006; Wood *et al.*, 2006; Sung & Amasino, 2004).

Whole genome mapping of H3K27me3 revealed its localization to roughly 15-20% of *Arabidopsis* genes, in particular, to the promoter and transcribed regions (Engelhorn *et al.*, 2017; Zhang *et al.*, 2007b). In some instances, H3K27me3 deposition is dependent on the epigenetic context of the target genes, such as presence of non-CG methylation or enrichment of TEs (Liu *et al.*, 2016; Moreno-Romero *et al.*, 2016; Zhou *et al.*, 2016; Dong *et al.*, 2012; Weinhofer *et al.*, 2010).

In *Arabidopsis*, LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), the plant homolog of animal HETEROCHROMATIN PROTEIN 1 (HP1), has been described as a “reader” of H3K27me3 (Turck *et al.*, 2007; Zhang *et al.*, 2007c). LHP1 recognizes and binds to Polycomb-deposited H3K27me3 along the key regulators of developmental transitions such as flowering and the differentiation from shoot meristems to leaves. Furthermore, LHP1 interacts with the PRC2 component MSI1 (Derkacheva *et al.*, 2013) and members of PRC1 (Xu & Shen, 2008), highlighting the connection between histone writers and readers in the regulation of the genome (Berry *et al.*, 2017; Merini *et al.*, 2017; Zhou *et al.*, 2017; Veluchamy *et al.*, 2016).

Methylation of arginine in *Arabidopsis* is mediated by protein arginine methyltransferases (PRMT) (Ahmad & Cao, 2012). Symmetrical dimethylation of H3R17 is catalyzed by AtPRMT4a and AtPRMT4b and reinforces silencing of *FLC* (Niu *et al.*, 2012). AtPRMT5 symmetrically dimethylates H4R3. The same residue is asymmetrically methylated by AtPRMT10, AtPRMT1a and AtPRMT1b, influencing *FLC* expression (Niu *et al.*, 2012; Schmitz *et al.*, 2008).

1.2.6 Histone lysine demethylases

The homeostasis of histone methylation is maintained by “erasers”, or histone lysine demethylases, which are broadly classified based on their structure into Lysine Specific Demethylases (LSD)-like (LSD) or Jumonji C (JmjC)-domain-containing protein family (JMJ) [Reviewed in (Trojer & Reinberg, 2006)]. Whereas histone lysine demethylases from both LSD (LSD, LSD1 and LSD2) and JMJ (JMJ14, JMJ15 and JMJ18) families remove H3K4 methylation (Lu *et al.*, 2010; Shi *et al.*, 2004), H3K36 methylation is removed by JMJ30 (Gan *et al.*, 2015).

Erasers of H3K27me1 have not yet been identified. Demethylation of H3K27me2 and H3K27me3 is mediated by demethylases from the JMJ family: EARLY FLOWERING 5 (ELF6/JMJ11), RELATIVE OF EARLY FLOWERING 6 (REF6/JMJ12), JMJ30 and JMJ32 (Gan *et al.*, 2014; Lu *et al.*, 2011; Noh *et al.*, 2004). All of these demethylases have a role in the control of *FLC*. ELF6/JMJ11, in particular, promotes the epigenetic reprogramming of *FLC* during gametogenesis (Crevillen *et al.*, 2014).

1.2.7 Histone acetylation: writers, readers and erasers

Histone acetylation is generally recognised as a marker for active transcription (Sequeira-Mendes *et al.*, 2014; Roudier *et al.*, 2011). Its negative charge neutralizes the positive charge of histone tails, decreasing the affinity of the DNA with the nucleosome and allowing the accessibility of proteins that actively promote transcription (Lusser *et al.*, 2001). On histone H3, the prime acetylated residues are lysine in positions 9, 14, 18 and 23. And in histone H4, lysine 5, 8, 12, 16 and 20 are acetylated (Fuchs *et al.*, 2006). Deposition of histone acetylation is catalysed by histone acetyl transferases (HATs) and removed by histone deacetylases (HDACs). In *Arabidopsis*, HATs are subdivided into four families (Lusser *et al.*, 2001):

- 1) GNAT superfamily (HAG)
- 2) MYST superfamily (HAM)
- 3) CREB-binding family (HAC)
- 4) TATA-binding protein-associated factor family (HAF)

Whereas histone deacetylases are classified into three families:

- 1) Homologues of yeast Reduced Potassium Deficiency 3 (RPD3)
- 2) HD-tuins (HDT)
- 3) Homologs of yeast silent information regulator 2 (Sir2)

Table 1 compiles the histone writers and erasers described so far in *Arabidopsis*.

Table 1. Histone modifications described so far in *Arabidopsis* and their corresponding histone writers and erasers (Maeji & Nishimura, 2018).

Modification	Writer	Eraser	Reference
H3K4me1	SDG25/ATXR7	FLD, LDL1, LDL2, JMJ14, JMJ15	(Yang <i>et al.</i> , 2012b; Lu <i>et al.</i> , 2010; Jiang <i>et al.</i> , 2009; Tamada <i>et al.</i> , 2009; Jiang <i>et al.</i> , 2007; Liu <i>et al.</i> , 2007)
H3K4me2	SDG2/ATXR3, SDG4/ASHR3, SDG25/ATXR7, SDG30/ATX2	FLD, LDL1, LDL2, JMJ14, JMJ15, JMJ18	(Yang <i>et al.</i> , 2012a; Yang <i>et al.</i> , 2012b; Lu <i>et al.</i> , 2010; Jiang <i>et al.</i> , 2009; Tamada <i>et al.</i> , 2009; Saleh <i>et al.</i> , 2008; Xu <i>et al.</i> , 2008; Jiang <i>et al.</i> , 2007; Liu <i>et al.</i> , 2007; Sutter <i>et al.</i> , 1978)
H3K4me3	SDG2/ATXR3, SDG4/ASHR3, SDG25/ATXR7, SDG26/ASHH1, SDG27/ATX1	JMJ14, JMJ15, JMJ18	(Yang <i>et al.</i> , 2012a; Yang <i>et al.</i> , 2012b; Berr <i>et al.</i> , 2010; Lu <i>et al.</i> , 2010; Tamada <i>et al.</i> , 2009; Cartagena <i>et al.</i> , 2008; Xu <i>et al.</i> , 2008)
H3K9me1 H3K9me2	SDG9/SUVH5, SDG23/SUVH6, SDG31/SUVR4	IBM1 IBM1	(Saze <i>et al.</i> , 2008; Ebbs & Bender, 2006; Grini <i>et al.</i> , 2006; Ebbs <i>et al.</i> , 2005)
H3K9me3	SDG33/KYP/SUVH	Unknown	(Ebbs & Bender, 2006; Grini <i>et al.</i> , 2006; Ebbs <i>et al.</i> , 2005)
H3K27me1	SDG15/ATXR5, SDG34/ATXR6		(Jacob <i>et al.</i> , 2010)
H3K27me2 H3K27me3	SDG1/CLF, SDG5/MEA, SDG10/SWN	JMJ11/ELF6, JMJ12/REF6, JMJ30, JMJ32	(Gan <i>et al.</i> , 2014; Lu <i>et al.</i> , 2011; Wang <i>et al.</i> , 2006; Noh <i>et al.</i> , 2004; Grossniklaus <i>et al.</i> , 1998; Goodrich <i>et al.</i> , 1997)

H3K36me1	SDG4/ASHR3	Unknown	(Cartagena <i>et al.</i> , 2008)
H3K36me2	SDG4/ASHR3, SDG8/ASHH2, SDG25/ATXR7	JMJ30	(Lu <i>et al.</i> , 2011; Berr <i>et al.</i> , 2009; Cartagena <i>et al.</i> , 2008; Xu <i>et al.</i> , 2008)
H3K36me3	SDG4/ASHR3, SDG8/ASHH2, SDG26/ASHH1		(Lu <i>et al.</i> , 2011; Cartagena <i>et al.</i> , 2008; Xu <i>et al.</i> , 2008)
H3R17me2	AtPRMT4a, AtPRMT4b	Unknown	(Niu <i>et al.</i> , 2008)
H4R3me2	AtPRMT1a, AtPRMT1b, AtPRMT5, AtPRMT10		(Niu <i>et al.</i> , 2008; Niu <i>et al.</i> , 2007; Pei <i>et al.</i> , 2007)
H3K9ac, H3K14ac H3K18ac, H3K23ac H4K5ac, H4K8ac H4K12ac, H4K16ac H4K20ac	HAG1/GCN5, HAG2, HAG3, MCC1, HAM1, HAM2, HAC1, HAC2, HAC4, HAC5, HAC12, HAF1, HAF2	HDA2, HDA5, HDA6/RPD3B, HDA7, HDA8, HDA9, HDA10, HDA14, HDA15, HDA17, HDA18, HDA19, HDT1/HD2A, HDT2/HD2B, HDT3, HDT4, SRT1, SRT2	(Wang <i>et al.</i> , 2014)

1.3 Chromatin assembly

To ensure proper inheritance of the genetic material, chromatin needs to be duplicated almost concomitantly to the DNA. Meaning that the canonical histones, some of which are freshly synthesized, are incorporated into new nucleosomes during the S phase of the cell division cycle. Two different mechanisms maintain the kinetics of DNA synthesis and chromatin formation. On one hand, the replication fork assists the re-assembly of nucleosomes by loading histones from the parental nucleosome on either one of the DNA daughter strands (Gasser *et al.*, 1996). On the other hand, in a separate reaction, newly synthesized histones H3 and H4 come together to form the H3-H4 dimer and are *de novo* assembled into (H3-H4)₂ tetramers to organize the central 70-80 bp of DNA from the daughter DNA strands [Reviewed in (Sauer *et al.*, 2018)]. Subsequently, the octamer is complemented by the chaperone-mediated deposition of either new or parental H2A-H2B dimers, which organize the peripheral 30-40 pb on either side of the tetramer (Jackson, 1990). Histone chaperones mediate the step-wise processes of disassembly and reassembly of nucleosomes. *In vitro* studies showed that Chromatin Assembly Factor 1 (CAF-1) serves as a histone chaperone assisting the assembly and recruitment of newly synthesized (H3-H4)₂ tetramers as the first step of replication dependent nucleosome assembly (Kaufman *et al.*, 1995). The formation of the complex between DNA and the (H3-H4)₂ tetramer, does not require ATP hydrolysis. It is driven by the high affinity between the tetramer and the DNA in a step-wise mechanisms guided by CAF-1 [Reviewed in (Sauer *et al.*, 2018)].

The function and heterotrimeric composition of CAF-1 is highly conserved. It was first identified *in vitro* for its ability to assemble the histone octamer (Stillman, 1986). Further purification from human cell nuclear extract showed its potential to load histones into newly formed nucleosomes in a DNA replication-dependent manner (Smith & Stillman, 1989). Mammalian CAF-1 is composed of the subunits p150, p60 and p48. In *S. cerevisiae* it is known as Chromatin Assembly Complex (CAC) and composed of Cac1, Cac2 and Cac3, whereas in *Arabidopsis* the three CAF-1 subunits are encoded by the p150 homolog *FASCIATA 1* (*FAS1*), the p60 homolog, *FASCIATA 2* (*FAS2*), and the p48 homolog *MULTICOPY SUPPRESSOR OF IRA 1* (*MSI1*) [Reviewed in (Loyola & Almouzni, 2004)]. p150, the largest

CAF-1 subunit and homolog of CAC1 and FAS1, interacts with the Proliferating Cell Nuclear Antigen (PCNA), prompting the recruitment of CAF-1 to couple the DNA replication process to chromatin reconstitution (Jackson, 1990). The middle subunits, Cac2, p60, and FAS2 contain seven conserved WD domains as tandem repeats. The same holds true for the smallest subunits Cac3, p48 and MSI1 (van Nocker & Ludwig, 2003). Additionally, and independently of CAF-1 function, MSI1 proteins associate with other proteins such as members of PRC2, regulators of flowering, retinoblastoma proteins and regulators of the E2F pathway (Hennig *et al.*, 2005; Hennig *et al.*, 2003; Kohler *et al.*, 2003a; van Nocker & Ludwig, 2003). Lack of CAF-1 in mammals is lethal (Takami *et al.*, 2007). Interestingly, *Saccharomyces cerevisiae* is robust to dysfunctional CAC: deletion of either of the three subunits has only a minor effect on growth (Kaufman *et al.*, 1995). Similarly, disruption of the *Arabidopsis* CAF-1 subunits causes altered phenotypes to varying degrees. *fas1* and *fas2* mutants in *Arabidopsis* are viable, but they develop disorganized meristems associated with aberrant expression of two key markers of meristems: *WUSCHEL* in the shoot and *SCARECROW* in the root (Kaya *et al.*, 2001). As a consequence, the mutant plants display impaired development of roots, stem, leaves and floral organs; have defects in the cell fate transitions (Costa & Shaw, 2006; Exner *et al.*, 2006) and regulation of the cell cycle (Ramirez-Parra & Gutierrez, 2007). Recent work from our lab revealed that the lack of a functional CAF-1 underlies hyper activation of the immune system, which is accompanied by an unrestricted activation of defence related genes (Mozgova *et al.*, 2015). Consistent with the role of CAF-1 in the replication-coupled maintenance of heterochromatin, the molecular phenotype of *fas2* mutants include the disorganization of heterochromatin and reactivation of loci associated to silenced chromatin (Huang *et al.*, 2010; Loyola *et al.*, 2009; Huang *et al.*, 2007; Houlard *et al.*, 2006; Schonrock *et al.*, 2006b). CAF-1 is also involved in the maintenance of telomeres and ribosomal DNA repeats (rDNA), revealed by the fact that *fas1* and *fas2* mutants are prone to severe genomic rearrangements that include telomere shortening and loss of rDNA. This particular molecular phenotype becomes more severe across generations, coinciding with loss of plant vigour, and leads to infertility in the sixth-to-eight generation (Mozgova *et al.*, 2010).

1.4 Methods to study chromatin biology

The establishment and maintenance of chromatin is the result of complex dynamical networks operating at different spatiotemporal scales. Constant refinement of molecular and biochemical methods to isolate chromatin, coupled with imaging and genomic technologies to evaluate it, make it possible to address and integrate the multiple layers of increasing chromatin complexity: from genomic elements to nucleosome occupancy, higher order chromatin organization, genome topology and association to the nuclear lamina.

A commonly encountered challenge in the biochemical study of plant chromatin is the low concentration of histones per cell. Additionally, vegetative tissues are enriched in charged polycarbohydrates that confound the isolation of basic histone proteins. These issues can be circumvented by using proliferative tissues such as inflorescences that contain small cells depleted of chloroplasts, allowing the abundant extraction of histone proteins (Mahrez & Hennig, 2018).

The interlinked nature of the chromatin structure and regulation makes it possible to cross-correlate independent datasets from different spatiotemporal snapshots to generate an integrated view of the epigenetic regulation throughout the lifetime of an organism. However, to reach sound biological conclusions, the reliable interpretation of NGS data requires a careful consideration of potential biological and experimental biases, as well as systematic artefacts present in many NGS chromatin and transcriptome profiles. Many of these can be alleviated by implementing refined bioinformatic and statistical methods. Still, good experimental practices in epigenome experiments require a thorough design including controls, replicates and proper computational pipelines.

In principle, the standard pipeline used to interrogate the chromatin consists of fragmentation, enrichment, amplification, sequencing and mapping back to the reference genome. Depending on the biological question, the pipeline can be tailored to identify a particular chromatin status. Assisted by endonuclease fragmentation, deoxyribonuclease I (DNase I) digestion followed by sequencing reveals regions of open chromatin; whereas micrococcal nuclease (MNase) digest followed by sequencing detects the location of well-positioned nucleosomes. The enrichment step in both of these techniques requires fragment-size selection, since the oligonucleotide DNA directly recovered after digestion is the main interrogated molecule. Chromatin

immunoprecipitation (ChIP)-seq reveals the binding site of transcription factors, the presence of histone variants in the nucleosome and the addition of post-translational modifications to histone tails. This is elegantly accomplished by sonicating the chromatin and isolating the desired chromatin mark with specific antibodies. The same principle can be used to address the binary methylation status of cytosines using methylation sensitive antibodies. When more resolution is desired, bisulphite sequencing is used to read the methylation status of the genome at single nucleotide resolution, revealing not only the sequence context of a given cytosine, but also a consensus of its fractional level of methylation [Reviewed in (Meyer & Liu, 2014)]. Finally, genome-wide positional and enrichment information of the different epigenetic components can be correlated with digital values of the expression of the genome, recovered by enrichment of polyadenylated transcripts followed by sequencing. The sensitivity of RNA enrichment methods has made it possible to also profile non-coding RNAs, particularly important in PTGS and RdDM pathways.

The correct handling of NGS datasets driven by concise biological questions will continue to push forward the development of experimental and computational techniques, allowing a better understanding of the multiple levels of genome regulation. The following sections describe in more detail the relevant molecular and biochemical techniques coupled to NGS to interrogate the chromatin and its transcriptional readout.

1.4.1 Profiles of nucleosome occupancy and chromatin accessibility

Digestion of the chromatin with MNase has been a pivotal technique for chromatin research. It allows the purification of nucleosomes and is used in conjunction with high-throughput sequencing to profile the genome-wide occupancy of well-positioned nucleosomes. MNase is an endo-exonuclease able to digest double and single stranded DNA. On the first digestion step of native chromatin, MNase cleaves the linker region as an endonuclease and releases chromatosomes. Next, the exonuclease activity digests the linker regions progressing in both directions from the initial cleavage site. Processivity of the exonucleolytic activity forces the nucleosome to slide along unprotected DNA, exposing additional nucleotides for digestion and further degradation of the nucleosomal DNA (Catez *et al.*, 2003). The

degree of chromatin compaction affects the kinetics of the MNase enzymatic digestion in several ways: accessible chromatin increases the rate of the initial endonucleolytic attack, whereas a stable nucleosome core prevents sliding of the octamer and decreases the rate of exonucleolytic digestion. Hence, the chromatin structure and the length of the enzymatic reaction determine the recovery of tri- di- and mononucleosomal DNA fragments (Postnikov & Bustin, 1999). The recovered population of nucleosomes is then sequenced and nucleosome occupancy is profiled by modelling the read accumulation patterns along the genome.

DNA sequence motifs recognized by transcriptional regulatory proteins control gene expression as *cis*-regulatory elements. In order to be physically accessible, they are commonly found in regions of open chromatin depleted of nucleosomes. DNase is an endonuclease able to cut double stranded DNA at nucleosome-free regions known as DNase I hypersensitive sites (DHSs) and commonly associated to *cis*-regulatory elements. DNase I digestion is done on intact nuclei. Released DNA fragments are then sequenced and similar to MNase-seq, chromatin accessibility is recovered by read counting patterns along the genome (Cumbie *et al.*, 2015).

1.4.2 Expanding the chromatin repertoire: Histone post-translational modifications and histone variants

Histone modifications are detected in either native or formaldehyde crosslinked chromatin. The chromatin is fragmented by MNase in native chromatin and by sonication in cross-linked chromatin. Specific antibodies are then used to immunoprecipitate the DNA fragments associated to nucleosomes carrying the histone PTM in question. Following enrichment, the crosslinking, if any, is reversed to release the DNA. The immunoprecipitated DNA can be identified either by microarray hybridization (ChIP-chip) or by high-throughput sequencing (ChIP-seq). The same principle is used to immunoprecipitate and profile histone variants.

The fact that different functional groups can be deposited at multiple amino acids of histone tails, suggests the existence of combinatorial configurations of histone PTMs and a defined chromatin expression. It was suggested that this concerted action of specific histone modifications could be read as a code to predict the transcriptional response (Jenuwein & Allis, 2001). Albeit appealing,

this concept has proven to be rather simplistic since the actual chromatin expression is the result of integrated interactions between histone modifications, histone variants, non-coding RNAs, scaffold proteins and many more factors which define chromatin territories (Sequeira-Mendes *et al.*, 2014; Roudier *et al.*, 2011). Regardless, the more histone modifications we know, the better our understanding of the chromatin. Mass spectrometry (MS) is commonly used for the biochemical detection of histone PTMs. The sensitivity of this method allows for the discovery of uncharacterized modifications. Coupling the discovery with high-throughput sequencing, allows the detailed genome-wide profile of the histone PTM distribution, providing insights for its role in chromatin configuration and transcriptional control.

1.4.3 DNA methylation landscapes

Upon recognition of DNA methylation as an epigenetic mark (Riggs, 1975), the use of methylation sensitive enzymes was introduced as a way to detect DNA methylation in mammals, particularly in the most prevalently methylated CG context (Bird, 1978; Bird & Southern, 1978). Nowadays, Nanopore® sequencing developed by Pacific Biosciences is able to detect methylated DNA nucleotides while reading the DNA molecule in question. This is possible by analysing the kinetics of the polymerase used for single molecule sequencing (Flusberg *et al.*, 2010). At modified nucleotides, the processivity of the polymerase is subtly paused. This triggers a change in the electrical current signals detected by the sequencing instrument, which allows to predict the likelihood of the presence of methylated residues (Simpson *et al.*, 2017). One of the main advantages of the direct detection of methylated residues is that it automatically discards biases associated to sample enrichment before library preparation. However, nanopore is still an expensive, and therefore not widely used sequencing technique. Simultaneous DNA sequencing and methylation detection has been demonstrated only in bacterial and human genomes (Gigante *et al.*, 2018).

Other traditional methods exist to address genome-wide methylation profiles. The methyl functional group in the 5th carbon can either be detected by ChIP using specific antibodies, or by bisulphite conversion of unmethylated cytosines followed by sequencing (BS-seq) [Reviewed in (Kurdyukov & Bullock, 2016)].

So far, bisulphite sequencing is considered as the “gold standard” method for DNA methylation studies. Bisulphite treatment of denatured DNA deaminates unmethylated cytosines to convert them into uracils that will be read as thymines during the sequencing step. Methylated cytosines are protected against bisulphite oxidation and will be read as cytosines. The efficiency of the bisulphite conversion reaction can be evaluated by reading the methylation status of a known DNA control with unmethylated cytosines. After the bisulphite treatment, all the cytosines in the control should be recovered as thymines. Plants have an internal methylation control, since the cytosines in the chloroplast genome are not methylated.

Bisulphite treated reads are then compared to the reference genome to detect the bisulphite deamination events displayed as thymines instead of the expected original cytosines. Methylation values are expressed as a fraction of the reads in which cytosines at a given position were detected as thymines by the total number of reads covering the cytosine in question. Because the DNA methylation value represents an average of the DNA molecules present at the time of the snapshot, homogeneity of the cell population should be taken into consideration to prevent confounding effects diluting or inflating the methylation values.

1.4.4 3D genome and intranuclear distribution

NGS is also used to capture long-range genomic interactions that provide insights to reconstruct the topology of the genome. Locus-specific chromatin conformation capture (3C) quantifies the number of interactions between chromosomal loci that are brought together by the three-dimensional structure of the genome. The frequency of observed interactions can be converted into relative distances to reconstruct the spatial organization of the chromatin within the nucleus. Building up on the principle of capturing the closeness of genomic loci, chromatin conformation capture is coupled to genome-wide assessments to expand the scope of discovery by simultaneously detecting all possible pairs of fragments (Bonev & Cavalli, 2016; Cavalli & Misteli, 2013). While 4C recovers “one-against-all” interactions, Hi-C stands for “all-against-all”. High-throughput sequencing data recovered from 3C, and higher, techniques is used to draw genome-wide interaction maps and the organization of the genome into topologically associating domains (TADs). TADs

correlate with epigenetic markers that define functional chromatin territories. While TADs were not identified in *Arabidopsis*, 3D interaction between heterochromatin and Polycomb repressed domains was shown in *Arabidopsis* (Liu & Weigel, 2015; Feng *et al.*, 2014; Grob *et al.*, 2014).

2 Aims of the study

First used as a metaphor of how gene regulation modulates development, Waddington's "epigenetic landscape" is now generally used to describe a series of fate choices made by a developing cell towards differentiation and cell lineage specification. The epigenetic landscape depicts a ball at the top of a waving surface. The pathways that the ball might take as it rolls down are determined by a series of branching hills and valleys representing stable cellular states and the barriers preventing uncontrolled changes of cell identity (Waddington, 1957). The advent of high-throughput technologies has made it possible to obtain genome-wide snapshots of the chromatin configuration and the epigenetic landscape in response to physiological and developmental stimuli. Using *Arabidopsis thaliana* as a model and a combination of high-throughput sequencing and bioinformatic techniques, this thesis aims to expand our understanding of the composition and distribution of the building blocks of the chromatin: nucleosomes, histone variants, histone PTMs, and modifications on the DNA; as well as the corresponding transcriptional response. The knowledge produced by this thesis will contribute to our understanding of the dynamic nature of the chromatin and the different configurations of the epigenomic landscape driving cell-fate decisions in plants.

3 Results and discussion

3.1 Expanding the histone language: H3K36ac and H3K23me1 are conserved histone modifications not previously described in plants

High resolution mass spectrometry-based techniques are normally used to uncover single and combinatorial configurations of histone variants and histone PTMs (Mahrez & Hennig, 2018). The implementation of such biochemical techniques to study histone modifications and histone variants is sometimes cumbersome when using vegetative tissues of *Arabidopsis*, since they contain a low concentration of histones. To circumvent this issue, proliferative tissues are used instead. In this study, we used inflorescences from *Brassica oleracea* (cauliflower), a close relative of *Arabidopsis*, to purify individual histones and subject them to high resolution mass spectrometry to unveil the histone PTMs present in their N-terminal tails. We were able to generate a map of acetylated and methylated residues of histone H3. Most of the histone modifications we recovered were consistent with previously described combinations in *Arabidopsis* (Zhang et al., 2007a; Garcia et al., 2004), suggesting high conservation of plant histone modifications. Additionally, we identified two previously undescribed modifications in the N-terminal tail of plant histone H3: acetylation in lysine 36, H3K36ac (Mahrez *et al.*, 2016) and monomethylation of lysine 23 H3K23me1 (Trejo-Arellano *et al.*, 2017). Both modifications are conserved in tobacco, wheat, rice and spruce, suggesting a conserved function.

3.1.1 H3K36ac is enriched in euchromatin and colocalizes with activating histone marks

To determine the genome-wide enrichment map of the newly identified histone modifications, we performed ChIP-seq. To control for differential nucleosome density, H3K36ac and H3K23me1 reads were normalized by their respective anti-H3 read counts. H3K36ac was found to be enriched in euchromatin and depleted from heterochromatin. Within the gene body, it was found in the first 500 bp downstream of the TSS, probably playing a role during the early phases of transcription. Levels of H3K36ac enrichment did not correlate with the transcriptional readout, suggesting that it acts as a binary indicator of transcription, with full acetylation present at different transcriptional levels. The peak of enrichment of H3K36ac co-localized with the presence of the H2A.Z histone variant. Whether H3K36ac promotes incorporation of H2A.Z or the other way around, remains to be established. Similarly, the activating histone mark H3K4me3 was also found towards the 5'-end of transcriptionally active genes. H3K36me2 and H3K36me3 on the other hand, peaked towards the 3'-end of the gene body. The gradient of H3K36ac, H2A.Z, H3K4me3, H3K36me2, and H3K36me3 along transcribed genes provides a spatial map of histone modifications along transcription units (Figure 3).

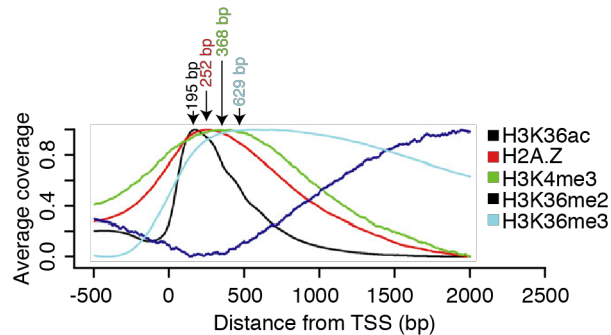


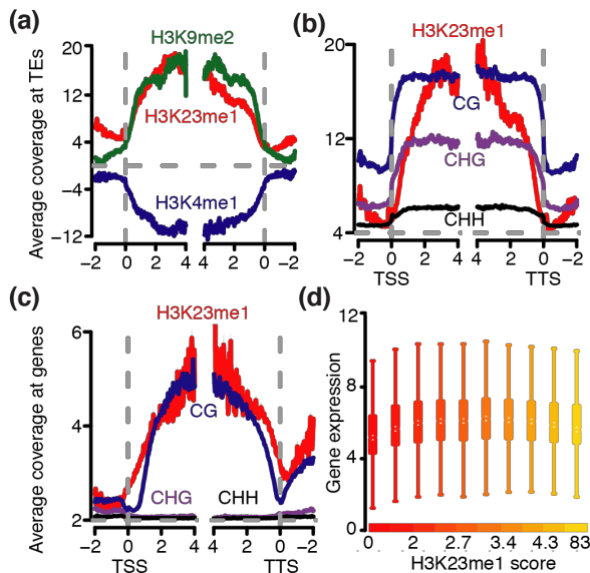
Figure 3. H3K36ac enrichment with respect to the TSS of genes. H3K36ac was found enriched downstream of the TSS of active genes, covering nucleosomes +1 and +2, where it colocalized with the H2A.Z histone variant. H3K36ac coexisted with other marks of active chromatin such as H3K4me1 and H3K36me3. This panel is an extract from the original figured first published in (Mahrez *et al.*, 2016).

We found that the homeostasis of H3K36ac is mediated by the HAT GCN5 and the HDAC encoded by *HDAC19*. Interestingly, our genome-wide data revealed a negative crosstalk between GCN5 and the HKMT SDG8 depositing H3K36me3. Disruption of GCN5 led to

higher levels of H3K36me3, and *sdg-8* mutants displayed gain of H3K36ac. This qualitative competition suggests that acetylation and methylation at H3K36 might act as a molecular switch.

3.1.2 H3K23me1 dually marks eu- and heterochromatin but does not have a direct effect in gene expression

H3K23me1 was preferentially found in heterochromatic chromocenters and pericentromeric TEs, where it coexisted with H3K9me2 and DNA methylation in the three contexts (Figure 4a-b). This is consistent with the previously described colocalization of H3K23me1 and the heterochromatic protein HP1 β in mammals (LeRoy *et al.*, 2012; Liu *et al.*, 2010b). Interestingly, we also found H3K23me1 along the coding region of transcriptionally active genes, where it colocalized with CG gbM (Figure 4c). However, presence and level of enrichment of H3K23me1 did not seem to have a direct effect on gene expression (Figure 4d).



*Figure 4. H3K23me1 enrichment along TEs and genes. (a) H3K23me1 colocalizes with the heterochromatic histone mark H3K9me2 (green) on TEs. Along those elements, H3K4me1 (blue) is generally excluded. (b) H3K23me1 was found together with TE methylation in the three contexts. (c) Whereas in genes it colocalized only with CG gene body methylation. (d) Presence and levels of H3K23me1 enrichment in the gene body did not seem to have a direct effect on gene expression. This figure was assembled from selected panels originally published in (Trejo-Arellano *et al.*, 2017).*

The homeostasis of H3K23me1 deposition required the undisrupted catalytic activity of SUVH4/KYP, the HKMT depositing H3K9me2, since *kyp* mutants displayed reduced levels of H3K23me1. Similarly, using Zebularine, a chemical blocker of DNA methylation, reduced H3K23me1 enrichment was observed at few analysed TEs. It remains

to be tested whether H3K23me1 can be recognised by Tudor-domain containing proteins to recruit methyltransferases, as previously shown for H3K9me2 in the reinforcement of silencing (Law *et al.*, 2013).

3.2 The *Arabidopsis* genome and chromatin configurations are robust to the lack of the replication dependent CAF-1

In *Arabidopsis*, the crucial role of CAF-1 in ensuring the correct transmission of chromatin states is made evident by the transgenerational aggravation of the developmental and molecular phenotypes upon the absence of a functional CAF-1 (Figure 5) (Li *et al.*, 2018; Mozgova *et al.*, 2018; Gouil & Baulcombe, 2016; Mozgova *et al.*, 2015; Mozgova *et al.*, 2010; Jacob *et al.*, 2009; Ortega-Galisteo *et al.*, 2008; Herr *et al.*, 2005). To further unveil the impact of a non-functional CAF-1 on the integrity of the genome across generations, we performed genomic resequencing of the sixth generation of self-fertilized *fas2* mutants (Muñoz-Viana *et al.*, 2017). We detected few genomic regions missing from selected locations. Particularly affected was chromosome 3 at the 45S rDNA and pericentromeric-silenced loci with high TE and rDNA density, as previously reported (Mozgova *et al.*, 2010). MNase-seq revealed that the nucleosome landscape of *Arabidopsis* is not severely affected by the lack of a functional CAF-1, suggesting that alternative nucleosome assemblers take over the function of CAF-1. Interestingly, promoters of defense-related genes were preferentially prone to lose nucleosomes. The increased accessibility at those particular genes suggests that upon CAF-1 disruption, the genome enters into a primed state that facilitates a quick response against external threats that compromise plant fitness. Together, these results reveal that the chromatin and nucleosome configuration of *Arabidopsis* is robust to the absence of the dedicated chromatin assembly factor CAF-1.



Figure 5. Transgenerational aggravation of *Arabidopsis* phenotype in *fas2* mutants. Plants lacking a functional CAF-1 display a progressive loss of plant vigour, from the second (G2), to the sixth generation (G6). The effect is visible as early as 13 days after germination, as displayed in the above pictures. Bars=5 mm. This panel is an extract from a figure first published in (Mozgova *et al.*, 2018).

3.3 The *Arabidopsis* methylome is robust to disrupted chromatin

There is increasing evidence supporting the regulatory feedbacks that exist between DNA methylation, histone PTMs and histone variants. Relevant examples show that the presence and level of enrichment of histone PTMs can recruit the DNA methylation machineries and reinforce silencing (Zhou *et al.*, 2018; Law *et al.*, 2013; Johnson *et al.*, 2008). DNA methylation competes with histone variants that promote transcription (Bewick *et al.*, 2016; Jarillo & Pineiro, 2015; Costas *et al.*, 2011; Zilberman *et al.*, 2008) and when present at the promoter of some genes, DNA methylation can even promote gene expression by still unknown mechanisms (Tang *et al.*, 2016; Lei *et al.*, 2015; Williams *et al.*, 2015).

Despite of the tight functional link between DNA methylation and the different components of the chromatin structure, we found that the methylation landscape of *Arabidopsis* is rather stable even in conditions that compromise the stability of the chromatin.

3.3.1 Transgenerational aggravation of CAF-1 mutants is accompanied by sequence-specific DNA methylation changes in *Arabidopsis*

Methylome alterations have been reported for mutants of the CAF-1 subunits (Stroud *et al.*, 2014; Pontvianne *et al.*, 2013). Given the discrete observed changes in the nucleosome configuration in the absence of a functional CAF-1 (Muñoz-Viana *et al.*, 2017) and the reported interdependent activity of DNA methyltransferases and chromatin remodelers (Zemach *et al.*, 2013; Kakutani *et al.*, 1996), we addressed the long-range changes in the methylome from the first (G1) and sixth (G6) generations of self-fertilized *fas2* mutants (Mozgova *et al.*, 2018). Globally, gene bodies were not affected in any methylation context, not even in G6 (Figure 6). In TEs, GC methylation remained stable from WT to G1 but increased in G6. CHG showed a stepwise transgenerational increment from G1 and G6, probably as a consequence of the reduced nucleosome occupancy, making the chromatin more accessible to CMT2 (Muñoz-Viana *et al.*, 2017; Zemach *et al.*, 2013). CHH context mildly decreased in G1 but was then heavily hypermethylated in G6. Detection of methylation changes at a higher resolution by detecting differentially methylated regions

(DMRs), revealed that most of the CG DMRs changed in G1 and remained stable towards G6. CHH DMRs behaved similarly. Most CHG DMRs were already hypermethylated in G1 and did not change in G6. The fact that TEs were CHG hypermethylated in the first generation of CAF-1 mutants, and that this methylation further increased towards the sixth generation, suggests that the transgenerational aggravation of the phenotype has an underlying epigenetic basis.

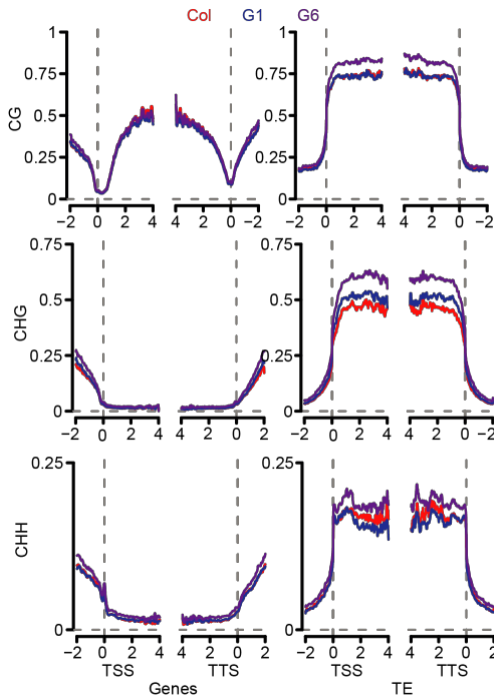


Figure 6. Context specific changes in the DNA methylation landscape of *fas2* mutants. DNA methylation along genes remained stable in all contexts. TEs however, displayed a different dynamic of change: CG remained stable in the first generation (G1) but then increased towards the sixth generation (G6). CHG was the only context displaying transgenerational progressive hypermethylation. CHH first decreased at G1. Towards G6, it recovered and became further hypermethylated. This panel is an extract from a figure first published in (Mozgova *et al.*, 2018).

3.3.2 The *Arabidopsis* methylome remains stable upon dark-induced senescence

Developmentally triggered senescence in *Arabidopsis* leaves displays a nuclear phenotype characterized by decondensed heterochromatin at chromocenters and relocation of H3K9me2 and H3K27me2. Euchromatin does not seem to be reconfigured to the same extent, but targeted H3K4me3 occurs at the promoter of *WRKY53*, which encodes for a transcription factor that regulates the expression of senescence activated genes (SAGs) (Ay *et al.*, 2009). We found that

decondensation of the chromocenters is a nuclear phenotype common to senescence, since we observed the same expansion of heterochromatin in interphase nuclei of senescent leaves induced by darkness as described in (Weaver & Amasino, 2001). Furthermore, we unveiled a massive downregulation of a transcriptional network in charge of maintaining the integrity of the chromatin and the epigenetic landscape (Figure 7). Members of this network included genes encoding for minichromosome maintenance proteins (MCM2-7), which act as checkpoints in the transition from G1 to S-phase of the cell cycle (Tuteja *et al.*, 2011; Shultz *et al.*, 2007). It furthermore included genes encoding for FAS1 and FAS2 subunits of CAF-1, a chaperone for (H3-H4)₂ assembly and required to preserve the chromatin structure during the replicative phase of the cell cycle (Kaya *et al.*, 2001; Smith & Stillman, 1989).

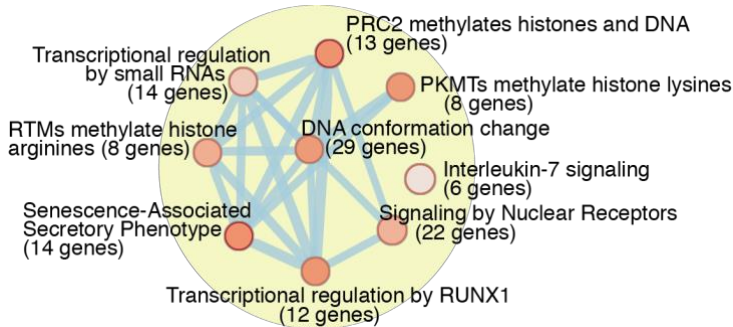


Figure 7. Global downregulation of genes encoding for members maintaining chromatin conformation and DNA methylation landscape.

The global downregulation of genes preserving the integrity of the chromatin was accompanied by deregulation of TEs, with young insertions being preferentially affected. Those changes however, were not accompanied by global disruption of the methylation landscape. In all three contexts, methylation levels remained stable along genes and TEs (Figure 8). Using 50 bp non-overlapping windows along the genome to detect differential methylation at higher DMRs resolution, we found that the CHH context was the most affected one, consistent with the observed downregulation of members of the CMT2/DDM1 and RdDM pathway. Global downregulation of the sentinels of the chromatin coupled with reactivation of TEs, suggests that upon senescence, fragmentation of the genome is promoted in order to recycle phosphate from the DNA. The few changes in the methylation

landscape at our sampling time suggest that a robust methylome in turn, might delay DNA degradation until the onset and last stages of senescence (Dodge, 1970; Butler, 1967).

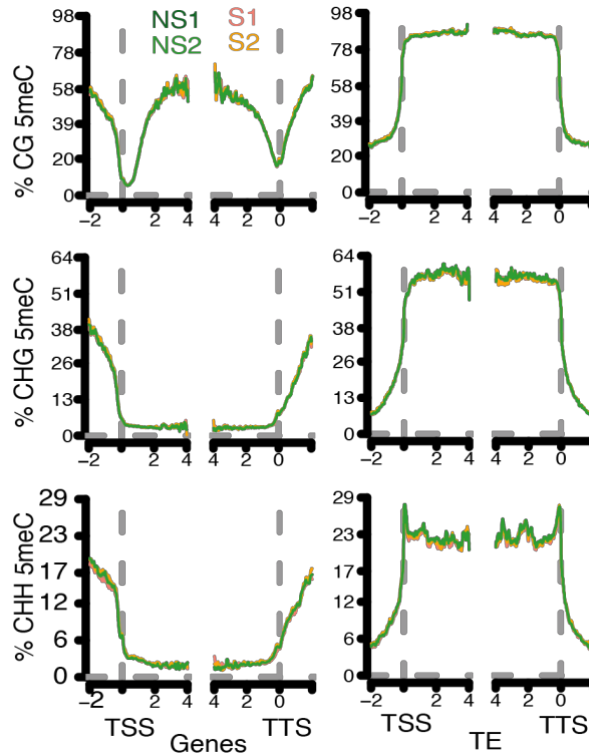


Figure 8. The methylome of dark-induced senescent leaves remained largely stable in the three methylation contexts. No pronounced changes were detected in the methylation levels along genes, nor along TEs.

4 Conclusions

The implementation of biochemical techniques to enrich for the components of the chromatin, coupled to high-throughput sequencing, allowed us to elucidate the genome-wide distribution of newly described histone modifications, and to reconstruct the dynamics of the *Arabidopsis* methylation landscape upon internal and external stimuli. Furthermore, cross-correlation of the data sets produced for this thesis with publicly available data, allowed us to contextualize our findings into the ever-growing knowledge produced by the chromatin and epigenetic research community.

H3K36ac was described for the first time in *Arabidopsis* as a conserved histone mark enriched in euchromatin, where it acts as a binary indicator of transcription. H3K23me1, also newly described by us, was found to be enriched in pericentromeric heterochromatin but also detectable along the coding region of expressed genes, where it coexists with CG gene body methylation.

Lack of a functional CAF-1 resulted in the depletion of repetitive and heterochromatic DNA from the genome and the reorganization of the nucleosome landscape to evict nucleosomes from the promoter of defense genes, causing priming of plants against pathogenic infections.

Importantly, my work revealed that aggravation of the CAF-1 mutant phenotype has an epigenetic basis. The DNA methylation landscape of CAF-1 was mostly affected towards the sixth generation of self-fertilized plants, with CHG methylation changes displaying a cumulative transgenerational disruption. This trend mirrored the phenotypic and molecular transgenerational aggravation of CAF-1 mutants, highlighting the importance of the correct configuration of the epigenome to execute the genetic information.

Lastly, my work revealed that the methylome is remarkably robust in the late phase of cellular reprogramming, as revealed by local, but no global changes of DNA methylation detected upon dark-induced senescence in *Arabidopsis* leaves. Despite an extensive decondensation of heterochromatic chromocenters, the deregulation of TEs and the severe downregulation of genes coding for scaffold proteins and maintainers of the chromatin, the dark-induced senescent methylome remained highly stable. Nevertheless, there were localized changes in CHH methylation that correlated with gene expression changes, indicating that senescence-induced CHH methylation changes have a regulatory role. Together, this data reveal that the terminal stage of plant life is accompanied by global changes in chromatin structure, but localized changes in DNA methylation, adding another example for the dynamics of DNA methylation during plant development.

5 Future perspectives

The genome-wide maps of histone modifications described in this thesis represent static spatiotemporal snapshots of the chromatin. To have a comprehensive understanding of their role in the regulation of gene expression, it is important to contextualize them into the landscape of chromatin territories and to unveil how are they collectively regulating the three-dimensional organization of the chromatin in the nuclear space.

Uncovering the mechanistic deposition and removal of the newly described histone modifications is of vital importance as well. We showed that GCN5 deposits H3K36ac but the HKMT catalyzing monomethylation of H3K23 remains to be discovered. Furthermore, the colocalization of H3K23me1 with the heterochromatic histone mark H3K9me2 suggests that they are both required for the correct maintenance of silenced loci. It is attractive to think that H3K23me1 might, similarly to H3K9me2, act as a tether to recruit DNA methyltransferases and reinforce silencing.

The surprising robustness of the nucleosome landscape, as well as the transgenerational reinforcement of CHG hypermethylation of TEs that accompany the progressive deterioration of the phenotype in CAF-1 mutants, highlight the importance of maintaining a methylome configuration that ensures correct inheritance of the genetic information. How are those DNA methylation patterns preserved? And why are they progressively hypermethylated? These are questions that remain to be answered.

The largely undisturbed methylation landscape upon dark-induced senescence in *Arabidopsis* leaves might represent a steric shield to modulate the uncontrolled fragmentation of the genome in the early stages of senescence. Characterization of the methylation landscape at later stages of either induced or developmentally programmed

senescence, might allow the reconstruction of the DNA methylation dynamics throughout the irreversible transition into a state of cellular arrest that culminates in cell death. This could be further extrapolated to the characterization of DNA methylation and histone modifications landscapes in tissues that are developmentally programmed to undergo cell death as part of their differentiation program. Since in principle, all the cells in the organism carry the same genetic information, one can only wonder what is the configuration of the epigenomic landscape in those cells that are programmed to undergo cell death in order to fulfil their function.

References

- Ahmad, A. & Cao, X. (2012). Plant PRMTs broaden the scope of arginine methylation. *J Genet Genomics*, 39(5), pp. 195-208.
- Allfrey, V.G. & Mirsky, A.E. (1964). Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science*, 144(3618), p. 559.
- Alvarez-Venegas, R., Pien, S., Sadler, M., Witmer, X., Grossniklaus, U. & Avramova, Z. (2003). ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. *Current Biology*, 13(8), pp. 627-37.
- Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G.D., Giaever, G. & Nislow, C. (2012). Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife*, 1, p. e00078.
- Amor, D.J., Bentley, K., Ryan, J., Perry, J., Wong, L., Slater, H. & Choo, K.H. (2004). Human centromere repositioning "in progress". *Proc Natl Acad Sci U S A*, 101(17), pp. 6542-7.
- Arents, G. & Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A*, 92(24), pp. 11170-4.
- Ausin, I., Mockler, T.C., Chory, J. & Jacobsen, S.E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. *Nat Struct Mol Biol*, 16(12), pp. 1325-7.
- Ay, N., Irmeler, K., Fischer, A., Uhlemann, R., Reuter, G. & Humbeck, K. (2009). Epigenetic programming via histone methylation at WRKY53 controls leaf senescence in Arabidopsis thaliana. *Plant J*, 58(2), pp. 333-46.
- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C. & Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824), pp. 120-124.
- Berr, A., McCallum, E.J., Alioua, A., Heintz, D., Heitz, T. & Shen, W.H. (2010). Arabidopsis histone methyltransferase SET DOMAIN GROUP8 mediates induction of the jasmonate/ethylene pathway genes in plant defense response to necrotrophic fungi. *Plant Physiol*, 154(3), pp. 1403-14.
- Berr, A., Xu, L., Gao, J., Cognat, V., Steinmetz, A., Dong, A. & Shen, W.H. (2009). SET DOMAIN GROUP25 encodes a histone methyltransferase and is involved in FLOWERING LOCUS C activation and repression of flowering. *Plant Physiol*, 151(3), pp. 1476-85.

- Berry, S., Rosa, S., Howard, M., Buhler, M. & Dean, C. (2017). Disruption of an RNA-binding hinge region abolishes LHP1-mediated epigenetic repression. *Genes Dev*, 31(21), pp. 2115-2120.
- Bewick, A.J., Ji, L., Niederhuth, C.E., Willing, E.M., Hofmeister, B.T., Shi, X., Wang, L., Lu, Z., Rohr, N.A., Hartwig, B., Kiefer, C., Deal, R.B., Schmutz, J., Grimwood, J., Stroud, H., Jacobsen, S.E., Schneeberger, K., Zhang, X. & Schmitz, R.J. (2016). On the origin and evolutionary consequences of gene body DNA methylation. *Proc Natl Acad Sci U S A*, 113(32), pp. 9111-6.
- Bewick, A.J. & Schmitz, R.J. (2017). Gene body DNA methylation in plants. *Curr Opin Plant Biol*, 36, pp. 103-110.
- Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R. & Lagrange, T. (2009). RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *Embo Reports*, 10, pp. 649-654.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev*, 16(1), pp. 6-21.
- Bird, A.P. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *J Mol Biol*, 118(1), pp. 49-60.
- Bird, A.P. & Southern, E.M. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis*. *J Mol Biol*, 118(1), pp. 27-47.
- Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., Froula, J., Kang, D.D., Malmstrom, R.R., Morgan, R.D., Posfai, J., Singh, K., Visel, A., Wetmore, K., Zhao, Z., Rubin, E.M., Korlach, J., Pennacchio, L.A. & Roberts, R.J. (2016). The Epigenomic Landscape of Prokaryotes. *PLoS Genetics*, 12(2), p. e1005854.
- Bonev, B. & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, 17(11), pp. 661-678.
- Bräutigam, K. & Cronk, Q. (2018). DNA Methylation and the Evolution of Developmental Complexity in Plants, 9(1447).
- Brink, R.A. (1956). A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. *Genetics*, 41(6), pp. 872-889.
- Butler, R.D. (1967). The Fine Structure of Senescing Cotyledons of Cucumber. *Journal of Experimental Botany*, 18(3), pp. 535-543.
- Calarco, J.P., Borges, F., Donoghue, M.T.A., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijo, J.A., Becker, J.D. & Martienssen, R.A. (2012). Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell*, 151(1), pp. 194-205.
- Cao, X. & Jacobsen, S.E. (2002a). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci U S A*, 99 Suppl 4, pp. 16491-8.
- Cao, X. & Jacobsen, S.E. (2002b). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Current Biology*, 12(13), pp. 1138-44.
- Caro, E., Stroud, H., Greenberg, M.V., Bernatavichute, Y.V., Feng, S., Groth, M., Vashisht, A.A., Wohlschlegel, J. & Jacobsen, S.E. (2012). The SET-domain protein SUV5 mediates

- H3K9me2 deposition and silencing at stimulus response genes in a DNA methylation-independent manner. *PLoS Genet*, 8(10), p. e1002995.
- Cartagena, J.A., Matsunaga, S., Seki, M., Kurihara, D., Yokoyama, M., Shinozaki, K., Fujimoto, S., Azumi, Y., Uchiyama, S. & Fukui, K. (2008). The Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. *Developmental Biology*, 315(2), pp. 355-368.
- Catez, F., Lim, J.H., Hock, R., Postnikov, Y.V. & Bustin, M. (2003). HMGN dynamics and chromatin function. *Biochem Cell Biol*, 81(3), pp. 113-22.
- Cavalli, G. & Misteli, T. (2013). Functional implications of genome topology. *Nat Struct Mol Biol*, 20(3), pp. 290-9.
- Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.H., Sung, Z.R. & Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development*, 131(21), pp. 5263-76.
- Charron, J.B., He, H., Elling, A.A. & Deng, X.W. (2009). Dynamic landscapes of four histone modifications during deetiolation in Arabidopsis. *Plant Cell*, 21(12), pp. 3732-48.
- Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S. & Peacock, W.J. (1997). Fertilization-independent seed development in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 94(8), pp. 4223-8.
- Cooper, D.W., Vandeber, J.I., Sharman, G.B. & Poole, W.E. (1971). Phosphoglycerate Kinase Polymorphism in Kangaroos Provides Further Evidence for Paternal-X Inactivation. *Nature-New Biology*, 230(13), pp. 155-&.
- Costa, S. & Shaw, P. (2006). Chromatin organization and cell fate switch respond to positional information in Arabidopsis. *Nature*, 439(7075), pp. 493-496.
- Costas, C., de la Paz Sanchez, M., Stroud, H., Yu, Y., Oliveros, J.C., Feng, S., Benguria, A., Lopez-Vidriero, I., Zhang, X., Solano, R., Jacobsen, S.E. & Gutierrez, C. (2011). Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. *Nat Struct Mol Biol*, 18(3), pp. 395-400.
- Crevillen, P., Yang, H., Cui, X., Greeff, C., Trick, M., Qiu, Q., Cao, X. & Dean, C. (2014). Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature*, 515(7528), pp. 587-90.
- Crouse, H.V. (1960). The Controlling Element in Sex Chromosome Behavior in Sciara. *Genetics*, 45(10), pp. 1429-1443.
- Cuerda-Gil, D. & Slotkin, R.K. (2016). Non-canonical RNA-directed DNA methylation. *Nat Plants*, 2(11), p. 16163.
- Cumbie, J.S., Filichkin, S.A. & Megraw, M. (2015). Improved DNase-seq protocol facilitates high resolution mapping of DNase I hypersensitive sites in roots in Arabidopsis thaliana. *Plant Methods*, 11, p. 42.
- Davey, C.A. & Richmond, T.J. (2002). DNA-dependent divalent cation binding in the nucleosome core particle. *Proc Natl Acad Sci U S A*, 99(17), pp. 11169-74.
- De Lucia, F., Crevillen, P., Jones, A.M., Greb, T. & Dean, C. (2008). A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc Natl Acad Sci U S A*, 105(44), pp. 16831-6.

- Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgova, I., Mahrez, W., Nanni, P., Bischof, S., Gruissem, W. & Hennig, L. (2013). Arabidopsis MSI1 connects LHP1 to PRC2 complexes. *EMBO J*, 32(14), pp. 2073-85.
- Dodge, J.D. (1970). Changes in Chloroplast Fine Structure During the Autumnal Senescence of *Betula* Leaves. *Annals of Botany*, 34(137), pp. 817-824.
- Doerfler, W. (2008). In pursuit of the first recognized epigenetic signal--DNA methylation: a 1976 to 2008 synopsis. *Epigenetics*, 3(3), pp. 125-33.
- Dong, X., Reimer, J., Gobel, U., Engelhorn, J., He, F., Schoof, H. & Turck, F. (2012). Natural variation of H3K27me3 distribution between two Arabidopsis accessions and its association with flanking transposable elements. *Genome Biol*, 13(12), p. R117.
- Du, J.M., Johnson, L.M., Groth, M., Feng, S.H., Hale, C.J., Li, S.S., Vashisht, A.A., Gallego-Bartolome, J., Wohlschlegel, J.A., Patel, D.J. & Jacobsen, S.E. (2014). Mechanism of DNA Methylation-Directed Histone Methylation by KRYPTONITE. *Molecular Cell*, 55(3), pp. 495-504.
- Duan, C.-G., Zhang, H., Tang, K., Zhu, X., Qian, W., Hou, Y.-J., Wang, B., Lang, Z., Zhao, Y., Wang, X., Wang, P., Zhou, J., Liang, G., Liu, N., Wang, C. & Zhu, J.-K. (2015). Specific but interdependent functions for Arabidopsis AGO4 and AGO6 in RNA-directed DNA methylation. *The EMBO journal*, 34(5), pp. 581-592.
- Duc, C., Benoit, M., Detourne, G., Simon, L., Poulet, A., Jung, M., Veluchamy, A., Latrasse, D., Le Goff, S., Cotterell, S., Tatout, C., Benhamed, M. & Probst, A.V. (2017). Arabidopsis ATRX Modulates H3.3 Occupancy and Fine-Tunes Gene Expression. *Plant Cell*, 29(7), pp. 1773-1793.
- Ebbs, M.L., Bartee, L. & Bender, J. (2005). H3 Lysine 9 Methylation Is Maintained on a Transcribed Inverted Repeat by Combined Action of SUVH6 and SUVH4 Methyltransferases. *Molecular and Cellular Biology*, 25(23), p. 10507.
- Ebbs, M.L. & Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell*, 18(5), pp. 1166-76.
- Elsaesser, S.J. & Allis, C.D. (2010). HIRA and Daxx constitute two independent histone H3.3-containing predeposition complexes. *Cold Spring Harb Symp Quant Biol*, 75, pp. 27-34.
- Engelhorn, J., Blanvillain, R., Kröner, C., Parrinello, H., Rohmer, M., Posé, D., Ott, F., Schmid, M. & Carles, C.C. (2017). Dynamics of H3K4me3 Chromatin Marks Prevails over H3K27me3 for Gene Regulation during Flower Morphogenesis in Arabidopsis thaliana, 1(2), p. 8.
- Enke, R.A., Dong, Z.C. & Bender, J. (2011). Small RNAs Prevent Transcription-Coupled Loss of Histone H3 Lysine 9 Methylation in Arabidopsis thaliana. *PLoS Genetics*, 7(10).
- Eun, C., Lorkovic, Z.J., Naumann, U., Long, Q., Havecker, E.R., Simon, S.A., Meyers, B.C., Matzke, A.J.M. & Matzke, M. (2011). AGO6 Functions in RNA-Mediated Transcriptional Gene Silencing in Shoot and Root Meristems in Arabidopsis thaliana. *PLoS One*, 6(10), p. e25730.
- Exner, V., Taranto, P., Schonrock, N., Gruissem, W. & Hennig, L. (2006). Chromatin assembly factor CAF-1 is required for cellular differentiation during plant development. *Development*, 133(21), pp. 4163-4172.

- Feng, S.H., Cokus, S.J., Schubert, V., Zhai, J.X., Pellegrini, M. & Jacobsen, S.E. (2014). Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Molecular Cell*, 55(5), pp. 694-707.
- Fincham, J.R.S. (1997). Epigenetic Mechanisms of Gene Regulation. Edited by V. E. A. Russo, R. A. Martienssen and A. D. Riggs. Cold Spring Harbor Laboratory Press, 1996. 693+xii pages. Price \$125. ISBN 0 87969 490 4. *Genetical Research*, 69(2), pp. 159-162.
- Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J. & Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, 7(6), pp. 461-U72.
- Fuchs, J., Demidov, D., Houben, A. & Schubert, I. (2006). Chromosomal histone modification patterns--from conservation to diversity. *Trends Plant Sci*, 11(4), pp. 199-208.
- Gan, E.S., Xu, Y., Wong, J.Y., Goh, J.G., Sun, B., Wee, W.Y., Huang, J. & Ito, T. (2014). Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in Arabidopsis. *Nat Commun*, 5, p. 5098.
- Gan, E.S., Xu, Y.F. & Ito, T. (2015). Dynamics of H3K27me3 methylation and demethylation in plant development. *Plant Signaling & Behavior*, 10(9).
- Garcia, B.A., Hunt, D.F., Shabanowitz, J., Johnson, L., Mollah, S., Jacobsen, S.E. & Muratore, T.L. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Research*, 32(22), pp. 6511-6518.
- Gasser, R., Koller, T. & Sogo, J.M. (1996). The stability of nucleosomes at the replication fork. *Journal of Molecular Biology*, 258(2), pp. 224-239.
- Gendall, A.R., Levy, Y.Y., Wilson, A. & Dean, C. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*, 107(4), pp. 525-35.
- Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C., Speed, T.P., Blewitt, M. & Ritchie, M. (2018). Using long-read sequencing to detect imprinted DNA methylation. *bioRxiv*.
- Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M. & Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. *Nature*, 386(6620), pp. 44-51.
- Gouil, Q. & Baulcombe, D.C. (2016). DNA Methylation Signatures of the Plant Chromomethyltransferases. *PLoS Genet*, 12(12), p. e1006526.
- Greb, T., Mylne, J.S., Crevillen, P., Geraldo, N., An, H.L., Gendall, A.R. & Dean, C. (2007). The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Current Biology*, 17(1), pp. 73-78.
- Greenberg, M.V.C., Ausin, I., Chan, S.W.L., Cokus, S.J., Cuperus, J.T., Feng, S.H., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C. & Jacobsen, S.E. (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics*, 6(3), pp. 344-354.
- Grini, P.E., Aalen, R.B., Sandvik, S.V., Johnsen, S.S., Thorstensen, T., Fischer, A. & Reuter, G. (2006). The Arabidopsis SUVH4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. *Nucleic Acids Research*, 34(19), pp. 5461-5470.
- Grob, S., Schmid, M.W. & Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Molecular Cell*, 55(5), pp. 678-693.

- Grossniklaus, U., Vielle-Calzada, J.P., Hoepfner, M.A. & Gagliano, W.B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science*, 280(5362), pp. 446-50.
- Guitton, A.E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.E., Grossniklaus, U. & Berger, F. (2004). Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. *Development*, 131(12), pp. 2971-81.
- Guo, L., Yu, Y., Law, J.A. & Zhang, X. (2010). SET DOMAIN GROUP2 is the major histone H3 lysine [corrected] 4 trimethyltransferase in Arabidopsis. *Proc Natl Acad Sci U S A*, 107(43), pp. 18557-62.
- He, X.J., Chen, T.P. & Zhu, J.K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Research*, 21(3), pp. 442-465.
- He, X.J., Hsu, Y.F., Zhu, S.H., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H.L. & Zhu, J.K. (2009). An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4-and RNA-Binding Protein. *Cell*, 137(3), pp. 498-508.
- Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet*, 9(1), pp. 15-26.
- Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R.T. (2018). Structure and function of archaeal histones. *PLoS Genetics*, 14(9).
- Hennig, L., Bouveret, R. & Grussem, W. (2005). MSI1-like proteins: an escort service for chromatin assembly and remodeling complexes. *Trends Cell Biol*, 15(6), pp. 295-302.
- Hennig, L., Taranto, P., Walser, M., Schonrock, N. & Grussem, W. (2003). Arabidopsis MSI1 is required for epigenetic maintenance of reproductive development. *Development*, 130(12), pp. 2555-65.
- Herr, A.J., Jensen, M.B., Dalmay, T. & Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science*, 308(5718), pp. 118-20.
- Holliday, R. & Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science*, 187(4173), pp. 226-32.
- Houben, A., Demidov, D., Gernand, D., Meister, A., Leach, C.R. & Schubert, I. (2003). Methylation of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA content. *Plant J*, 33(6), pp. 967-73.
- Houlard, M., Berlivet, S., Probst, A.V., Quivy, J.P., Hery, P., Almouzni, G. & Gerard, M. (2006). CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. *PLoS Genet*, 2(11), p. e181.
- Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L. & Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. *Science*, 324(5933), pp. 1451-4.
- Hu, L., Li, N., Xu, C., Zhong, S., Lin, X., Yang, J., Zhou, T., Yuliang, A., Wu, Y., Chen, Y.R., Cao, X., Zemach, A., Rustgi, S., von Wettstein, D. & Liu, B. (2014). Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc Natl Acad Sci U S A*, 111(29), pp. 10642-7.

- Huang, H., Yu, Z., Zhang, S., Liang, X., Chen, J., Li, C., Ma, J. & Jiao, R. (2010). Drosophila CAF-1 regulates HP1-mediated epigenetic silencing and pericentric heterochromatin stability. *J Cell Sci*, 123(Pt 16), pp. 2853-61.
- Huang, S., Zhou, H., Tarara, J. & Zhang, Z. (2007). A novel role for histone chaperones CAF-1 and Rtt106p in heterochromatin silencing. *EMBO J*, 26(9), pp. 2274-83.
- Huetzel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A.J.M. & Matzke, M. (2006). Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *Embo Journal*, 25(12), pp. 2828-2836.
- Huh, J.H., Bauer, M.J., Hsieh, T.F. & Fischer, R.L. (2008). Cellular programming of plant gene imprinting. *Cell*, 132(5), pp. 735-44.
- Ibarra, C.A., Feng, X.Q., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H. & Zilberman, D. (2012). Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. *Science*, 337(6100), pp. 1360-1364.
- Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I. & Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, 472(7341), pp. 115-U151.
- Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., Cheng, X., Schubert, I., Jenuwein, T. & Jacobsen, S.E. (2004). Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. *Chromosoma*, 112(6), pp. 308-15.
- Jackson, J.P., Lindroth, A.M., Cao, X. & Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature*, 416(6880), pp. 556-60.
- Jackson, V. (1990). In vivo studies on the dynamics of histone-DNA interaction: evidence for nucleosome dissolution during replication and transcription and a low level of dissolution independent of both. *Biochemistry*, 29(3), pp. 719-31.
- Jacob, Y., Bergamin, E., Donoghue, M.T.A., Mongeon, V., LeBlanc, C., Voigt, P., Underwood, C.J., Brunzelle, J.S., Michaels, S.D., Reinberg, D., Couture, J.F. & Martienssen, R.A. (2014). Selective Methylation of Histone H3 Variant H3.1 Regulates Heterochromatin Replication. *Science*, 343(6176), pp. 1249-1253.
- Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y.V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E. & Michaels, S.D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol*, 16(7), pp. 763-8.
- Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., Hassel, C., Gutierrez, C., Michaels, S.D. & Jacobsen, S.E. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature*, 466(7309), pp. 987-91.
- Jarillo, J.A. & Pineiro, M. (2015). H2A.Z mediates different aspects of chromatin function and modulates flowering responses in Arabidopsis. *Plant J*, 83(1), pp. 96-109.
- Jenuwein, T. & Allis, C.D. (2001). Translating the histone code. *Science*, 293(5532), pp. 1074-80.
- Jiang, D., Gu, X. & He, Y. (2009). Establishment of the Winter-Annual Growth Habit via FRIGIDA-Mediated Histone Methylation at

- <FLOWERING LOCUS C in <Arabidopsis. *The Plant Cell*, 21(6), p. 1733.
- Jiang, D., Wang, Y., Wang, Y. & He, Y. (2008). Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components. *PLoS One*, 3(10), p. e3404.
- Jiang, D., Yang, W., He, Y. & Amasino, R.M. (2007). <Arabidopsis Relatives of the Human Lysine-Specific Demethylase1 Repress the Expression of <FWA and <FLOWERING LOCUS C and Thus Promote the Floral Transition. *The Plant Cell*, 19(10), p. 2975.
- Johnson, L.M., Du, J.M., Hale, C.J., Bischof, S., Feng, S.H., Chodavarapu, R.K., Zhong, X.H., Marson, G., Pellegrini, M., Segal, D.J., Patel, D.J. & Jacobsen, S.E. (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature*, 507(7490), pp. 124-+.
- Johnson, L.M., Law, J.A., Khattar, A., Henderson, I.R. & Jacobsen, S.E. (2008). SRA-Domain Proteins Required for DRM2-Mediated De Novo DNA Methylation. *PLoS Genetics*, 4(11).
- Jones, L., Hamilton, A.J., Voinnet, O., Thomas, C.L., Maule, A.J. & Baulcombe, D.C. (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell*, 11(12), pp. 2291-301.
- Jones, R.N. (2005). McClintock's controlling elements: the full story. *Cytogenetic and Genome Research*, 109(1-3), pp. 90-103.
- Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K. & Richards, E.J. (1996). Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proceedings of the National Academy of Sciences*, 93, pp. 12406-12411.
- Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M. & Matzke, A.J. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Current Biology*, 14(9), pp. 801-5.
- Kato, M., Miura, A., Bender, J., Jacobsen, S.E. & Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. *Current Biology*, 13(5), pp. 421-6.
- Kaufman, P.D., Kobayashi, R., Kessler, N. & Stillman, B. (1995). The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell*, 81(7), pp. 1105-14.
- Kawakatsu, T., Huang, S.C., Jupe, F., Sasaki, E., Schmitz, R.J., Urlich, M.A., Castanon, R., Nery, J.R., Barragan, C., He, Y., Chen, H., Dubin, M., Lee, C.R., Wang, C., Bemm, F., Becker, C., O'Neil, R., O'Malley, R.C., Quarless, D.X., Genomes, C., Schork, N.J., Weigel, D., Nordborg, M. & Ecker, J.R. (2016). Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. *Cell*, 166(2), pp. 492-505.
- Kaya, H., Shibahara, K.I., Taoka, K.I., Iwabuchi, M., Stillman, B. & Araki, T. (2001). FASCIATA genes for chromatin assembly factor-I in arabidopsis maintain the cellular organization of apical meristems. *Cell*, 104(1), pp. 131-42.
- Kim, G.T., Tsukaya, H. & Uchimiya, H. (1998). The CURLY LEAF gene controls both division and elongation of cells during the expansion of the leaf blade in Arabidopsis thaliana. *Planta*, 206(2), pp. 175-83.

- Kohler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U. & Grissem, W. (2003a). Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO J*, 22(18), pp. 4804-14.
- Kohler, C., Hennig, L., Spillane, C., Pien, S., Grissem, W. & Grossniklaus, U. (2003b). The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev*, 17(12), pp. 1540-53.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4), pp. 693-705.
- Kurdyukov, S. & Bullock, M. (2016). DNA Methylation Analysis: Choosing the Right Method. *Biology (Basel)*, 5(1).
- Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A. & Jacobsen, S.E. (2010). A Protein Complex Required for Polymerase V Transcripts and RNA-Directed DNA Methylation in Arabidopsis. *Current Biology*, 20(10), pp. 951-956.
- Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M., Strahl, B.D., Patel, D.J. & Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature*, 498(7454), pp. 385-9.
- Law, J.A. & Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3), pp. 204-220.
- Lei, M.G., Zhang, H.M., Julian, R., Tang, K., Xie, S.J. & Zhu, J.K. (2015). Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11), pp. 3553-3557.
- LeRoy, G., Chepelev, I., DiMaggio, P.A., Blanco, M.A., Zee, B.M., Zhao, K. & Garcia, B.A. (2012). Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. *Genome Biology*, 13(8), p. R68.
- Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C. & Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A*, 107(32), pp. 14075-80.
- Li, Y., Kumar, S. & Qian, W.Q. (2018). Active DNA demethylation: mechanism and role in plant development. *Plant Cell Reports*, 37(1), pp. 77-85.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. & Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, 133(3), pp. 523-36.
- Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B. & Ecker, J.R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), pp. 315-22.
- Liu, C., Lu, F., Cui, X. & Cao, X. (2010a). Histone Methylation in Higher Plants, 61(1), pp. 395-420.
- Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M. & Weigel, D. (2016). Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res*, 26(8), pp. 1057-68.
- Liu, C. & Weigel, D. (2015). Chromatin in 3D: progress and prospects for plants. *Genome Biology*, 16.

- Liu, H., Galka, M., Iberg, A., Wang, Z., Li, L., Voss, C., Jiang, X., Lajoie, G., Huang, Z., Bedford, M.T. & Li, S.S.C. (2010b). Systematic Identification of Methyllysine-Driven Interactions for Histone and Nonhistone Targets. *Journal of Proteome Research*, 9(11), pp. 5827-5836.
- Liu, X., Yu, C.W., Duan, J., Luo, M., Wang, K., Tian, G., Cui, Y. & Wu, K. (2012). HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in Arabidopsis. *Plant Physiol*, 158(1), pp. 119-29.
- Liu, Y., Koornneef, M. & Soppe, W.J.J. (2007). The Absence of Histone H2B Monoubiquitination in the *Arabidopsis* *hub1* (hub1) Mutant Reveals a Role for Chromatin Remodeling in Seed Dormancy. *The Plant Cell*, 19(2), p. 433.
- Liu, Z.W., Shao, C.R., Zhang, C.J., Zhou, J.X., Zhang, S.W., Li, L., Chen, S., Huang, H.W., Cai, T. & He, X.J. (2014). The SET Domain Proteins SUVH2 and SUVH9 Are Required for Pol V Occupancy at RNA-Directed DNA Methylation Loci. *PLoS Genetics*, 10(1).
- Loidl, P. (2004). A plant dialect of the histone language. *Trends Plant Sci*, 9(2), pp. 84-90.
- Loyola, A. & Almouzni, G. (2004). Histone chaperones, a supporting role in the limelight. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1677(1), pp. 3-11.
- Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J.P., Imhof, A., Nakatani, Y., Dent, S.Y. & Almouzni, G. (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *Embo Reports*, 10(7), pp. 769-75.
- Lu, F., Cui, X., Zhang, S., Liu, C. & Cao, X. (2010). JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis. *Cell Research*, 20, p. 387.
- Lu, S.X., Knowles, S.M., Webb, C.J., Celaya, R.B., Cha, C., Siu, J.P. & Tobin, E.M. (2011). The Jumonji C domain-containing protein JMJ30 regulates period length in the Arabidopsis circadian clock. *Plant Physiol*, 155(2), pp. 906-15.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), pp. 251-60.
- Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J. & Chaudhury, A.M. (1999). Genes controlling fertilization-independent seed development in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 96(1), pp. 296-301.
- Lusser, A., Kolle, D. & Loidl, P. (2001). Histone acetylation: lessons from the plant kingdom. *Trends Plant Sci*, 6(2), pp. 59-65.
- Lyon, M.F. (1961). Gene Action in X-Chromosome of Mouse (*Mus Musculus* L). *Nature*, 190(477), pp. 372-&.
- Maeji, H. & Nishimura, T. (2018). Epigenetic Mechanisms in Plants. In.
- Mahrez, W., Arellano, M.S.T., Moreno-Romero, J., Nakamura, M., Shu, H., Nanni, P., Köhler, C., Grissem, W. & Hennig, L. (2016). H3K36ac Is an Evolutionary Conserved Plant Histone Modification That Marks Active Genes. *Plant Physiology*, 170, pp. 1566-1577.
- Mahrez, W. & Hennig, L. (2018). Mapping of Histone Modifications in Plants by Tandem Mass Spectrometry. *Methods Mol Biol*, 1675, pp. 131-145.
- Malagnac, F., Bartee, L. & Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J*, 21(24), pp. 6842-52.
- Mari-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V. & Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. *Nat Genet*, 45(9), pp. 1029-39.

- Matzke, M.A., Kanno, T. & Matzke, A.J. (2015). RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu Rev Plant Biol*, 66, pp. 243-67.
- McClintock, B. (1956). Controlling Elements and the Gene, 21, pp. 197-216.
- McCue, A.D., Panda, K., Nuthikattu, S., Choudury, S.G., Thomas, E.N. & Slotkin, R.K. (2015). ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *Embo Journal*, 34(1), pp. 20-35.
- Merini, W., Romero-Campero, F.J., Gomez-Zambrano, A., Zhou, Y., Turck, F. & Calonje, M. (2017). The Arabidopsis Polycomb Repressive Complex 1 (PRC1) Components AtBMI1A, B, and C Impact Gene Networks throughout All Stages of Plant Development. *Plant Physiol*, 173(1), pp. 627-641.
- Meyer, C.A. & Liu, X.S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev Genet*, 15(11), pp. 709-21.
- Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J. & Mathieu, O. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature*, 461(7262), pp. 427-30.
- Moreno-Romero, J., Jiang, H., Santos-Gonzalez, J. & Kohler, C. (2016). Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. *EMBO J*, 35(12), pp. 1298-311.
- Mozgova, I. & Hennig, L. (2015). The polycomb group protein regulatory network. *Annu Rev Plant Biol*, 66, pp. 269-96.
- Mozgova, I., Mokros, P. & Fajkus, J. (2010). Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45S rDNA in Arabidopsis thaliana. *Plant Cell*, 22(8), pp. 2768-80.
- Mozgova, I., Wildhaber, T., Liu, Q., Abou-Mansour, E., L'Haridon, F., Metraux, J.P., Grissem, W., Hofius, D. & Hennig, L. (2015). Chromatin assembly factor CAF-1 represses priming of plant defence response genes. *Nat Plants*, 1, p. 15127.
- Mozgova, I., Wildhaber, T., Trejo-Arellano, M.S., Fajkus, J., Roszak, P., Köhler, C. & Hennig, L. (2018). Transgenerational phenotype aggravation in CAF-1 mutants reveals parent-of-origin specific epigenetic inheritance, 220(3), pp. 908-921.
- Muñoz-Viana, R., Wildhaber, T., Trejo-Arellano, M.S., Mozgová, I. & Hennig, L. (2017). Arabidopsis Chromatin Assembly Factor 1 is required for occupancy and position of a subset of nucleosomes, 92(3), pp. 363-374.
- Muyle, A. & Gaut, B.S. (2018). Loss of gene body methylation in *Eutrema salsugineum* is associated with reduced gene expression. *Mol Biol Evol*.
- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F. & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, 543(7643), pp. 72-+.
- Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A., Burke, J.M., Udall, J.A., Egesi, C., Schmutz, J., Grimwood, J., Jackson, S.A., Springer, N.M. & Schmitz, R.J. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biol*, 17(1), p. 194.
- Niu, L., Lu, F., Pei, Y., Liu, C. & Cao, X. (2007). Regulation of flowering time by the protein arginine methyltransferase AtPRMT10. *Embo Reports*, 8(12), p. 1190.

- Niu, L., Lu, F., Zhao, T., Liu, C. & Cao, X. (2012). The enzymatic activity of Arabidopsis protein arginine methyltransferase 10 is essential for flowering time regulation. *Protein Cell*, 3(6), pp. 450-9.
- Niu, L., Zhang, Y., Pei, Y., Liu, C. & Cao, X. (2008). Redundant Requirement for a Pair of PROTEIN ARGININE METHYLTRANSFERASE4 Homologs for the Proper Regulation of Arabidopsis Flowering Time. *Plant Physiology*, 148(1), p. 490.
- Noh, B., Lee, S.H., Kim, H.J., Yi, G., Shin, E.A., Lee, M., Jung, K.J., Doyle, M.R., Amasino, R.M. & Noh, Y.S. (2004). Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription factor proteins in the regulation of Arabidopsis flowering time. *Plant Cell*, 16(10), pp. 2601-13.
- Noyer-Weidner, M. & Trautner, T.A. (1993). Methylation of DNA in Prokaryotes. In: Jost, J.-P. & Saluz, H.-P. (eds) *DNA Methylation: Molecular Biology and Biological Significance*. Basel: Birkhäuser Basel, pp. 39-108. Available from: https://doi.org/10.1007/978-3-0348-9118-9_4.
- Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N. & Slotkin, R.K. (2013). The Initiation of Epigenetic Silencing of Active Transposable Elements Is Triggered by RDR6 and 21-22 Nucleotide Small Interfering RNAs. *Plant Physiology*, 162(1), pp. 116-131.
- Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P. & Fischer, R.L. (1996). A mutation that allows endosperm development without fertilization. *Proc Natl Acad Sci U S A*, 93(11), pp. 5319-24.
- Ohno, S., Kaplan, W.D. & Kinoshita, R. (1959). Formation of the Sex Chromatin by a Single X-Chromosome in Liver Cells of Rattus-Norvegicus. *Experimental Cell Research*, 18(2), pp. 415-418.
- Ortega-Galisteo, A.P., Morales-Ruiz, T., Ariza, R.R. & Roldan-Arjona, T. (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Molecular Biology*, 67(6), pp. 671-681.
- Pei, Y., Niu, L., Lu, F., Liu, C., Zhai, J., Kong, X. & Cao, X. (2007). Mutations in the Type II Protein Arginine Methyltransferase AtPRMT5 Result in Pleiotropic Developmental Defects in Arabidopsis. *Plant Physiology*, 144(4), p. 1913.
- Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S. & Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. *Proc Natl Acad Sci U S A*, 104(16), pp. 6752-7.
- Peterson, C.L. & Herskowitz, I. (1992). Characterization of the Yeast Swi1, Swi2, and Swi3 Genes, Which Encode a Global Activator of Transcription. *Cell*, 68(3), pp. 573-583.
- Pien, S., Fleury, D., Mylne, J.S., Crevillen, P., Inze, D., Avramova, Z., Dean, C. & Grossniklaus, U. (2008). ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation. *Plant Cell*, 20(3), pp. 580-8.
- Pontvianne, F., Blevins, T., Chandrasekhara, C., Mozgová, I., Hassel, C., Pontes, O.M.F., Tucker, S., Mokroš, P., Muchová, V., Fajkus, J. & Pikaard, C.S. (2013). Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes & Development*, 27(14), pp. 1545-1550.
- Pontvianne, F., Blevins, T. & Pikaard, C.S. (2010). Arabidopsis Histone Lysine Methyltransferases. In: (Advances in Botanical Research, pp. 1-22.

- Postnikov, Y.V. & Bustin, M. (1999). Analysis of HMG-14/-17-containing chromatin. *Methods Mol Biol*, 119, pp. 303-10.
- Pusarla, R.H. & Bhargava, P. (2005). Histones in functional diversification - Core histone variants. *Febs Journal*, 272(20), pp. 5149-5168.
- Ramirez-Parra, E. & Gutierrez, C. (2007). The many faces of chromatin assembly factor 1. *Trends Plant Sci*, 12(12), pp. 570-6.
- Raynaud, C., Sozzani, R., Glab, N., Domenichini, S., Perennes, C., Cella, R., Kondorosi, E. & Bergounioux, C. (2006). Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. *Plant J*, 47(3), pp. 395-407.
- Reeve, J.N., Sandman, K. & Daniels, C.J. (1997). Archaeal histones, nucleosomes, and transcription initiation. *Cell*, 89(7), pp. 999-1002.
- Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet*, 14(1), pp. 9-25.
- Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Despres, B., Drevensek, S., Barneche, F., Derozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, M.L., Robin, S., Caboche, M. & Colot, V. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J*, 30(10), pp. 1928-38.
- Rowley, M.J., Rothi, M.H., Bohmdorfer, G., Kucinski, J. & Wierzbicki, A.T. (2017). Long-range control of gene expression via RNA-directed DNA methylation. *PLoS Genetics*, 13(5).
- Saleh, A., Alvarez-Venegas, R., Yilmaz, M., Le, O., Hou, G., Saddler, M., Al-Abdallat, A., Xia, Y., Lu, G., Ladunga, I. & Avramova, Z. (2008). The Highly Similar Arabidopsis Homologs of Trithorax ATX1 and ATX2 Encode Proteins with Divergent Biochemical Functions. *The Plant Cell*, 20(3), p. 568.
- Sauer, P.V., Panne, D., Gu, Y., Luger, K., Liu, W.H., Mattioli, F. & Churchill, M.E.A. (2018). Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Research*, 46(19), pp. 9907-9917.
- Saze, H., Shiraishi, A., Miura, A. & Kakutani, T. (2008). Control of Genic DNA Methylation by a jmjC Domain-Containing Protein in Arabidopsis thaliana. *Science*, 319(5862), p. 462.
- Schmitz, R.J., Sung, S. & Amasino, R.M. (2008). Histone arginine methylation is required for vernalization-induced epigenetic silencing of FLC in winter-annual Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 105(2), pp. 411-6.
- Schonrock, N., Bouveret, R., Leroy, O., Borghi, L., Kohler, C., Gruissem, W. & Hennig, L. (2006a). Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway. *Genes Dev*, 20(12), pp. 1667-78.
- Schonrock, N., Exner, V., Probst, A., Gruissem, W. & Hennig, L. (2006b). Functional genomic analysis of CAF-1 mutants in Arabidopsis thaliana. *J Biol Chem*, 281(14), pp. 9560-8.
- Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T. & Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J*, 25(19), pp. 4638-49.
- Sequeira-Mendes, J., Araguez, I., Peiro, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U. & Gutierrez, C. (2014). The Functional Topography of the Arabidopsis Genome

- Is Organized in a Reduced Number of Linear Motifs of Chromatin States. *Plant Cell*, 26(6), pp. 2351-2366.
- Serra-Cardona, A. & Zhang, Z. (2018). Replication-Coupled Nucleosome Assembly in the Passage of Epigenetic Information and Cell Identity. *Trends Biochem Sci*, 43(2), pp. 136-148.
- Shi, Y.J., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A. & Shi, Y. (2004). Histone demethylation mediated by the nuclear arnine oxidase homolog LSD1. *Cell*, 119(7), pp. 941-953.
- Shu, H., Nakamura, M., Siretskiy, A., Borghi, L., Moraes, I., Wildhaber, T., Gruijssem, W. & Hennig, L. (2014). Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes. *Genome Biol*, 15(4), p. R62.
- Shultz, R.W., Tatini, V.M., Hanley-Bowdoin, L. & Thompson, W.F. (2007). Genome-Wide Analysis of the Core DNA Replication Machinery in the Higher Plants Arabidopsis and Rice. *Plant Physiology*, 144, pp. 1697-1714.
- Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J. & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods*, 14(4), pp. 407-410.
- Simpson, R.T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*, 17(25), pp. 5524-31.
- Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A. & Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, 136(3), pp. 461-72.
- Smith, S. & Stillman, B. (1989). Purification and Characterization of Caf-I, a Human Cell Factor Required for Chromatin Assembly during DNA-Replication *In vitro*. *Cell*, 58(1), pp. 15-25.
- Stedman, E. & Stedman, E. (1950). Cell Specificity of Histones. *Nature*, 166(4227), pp. 780-781.
- Stillman, B. (1986). Chromatin assembly during SV40 DNA replication *in vitro*. *Cell*, 45(4), pp. 555-65.
- Strahl, B.D. & Allis, C.D. (2000). The language of covalent histone modifications. *Nature*, 403(6765), pp. 41-5.
- Stroud, H., Do, T., Du, J.M., Zhong, X.H., Feng, S.H., Johnson, L., Patel, D.J. & Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature Structural & Molecular Biology*, 21(1), pp. 64-+.
- Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V. & Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*, 152(1-2), pp. 352-64.
- Stroud, H., Otero, S., Desvoyes, B., Ramirez-Parra, E., Jacobsen, S.E. & Gutierrez, C. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 109(14), pp. 5370-5.
- Sung, S.B. & Amasino, R.M. (2004). Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. *Nature*, 427(6970), pp. 159-164.
- Sung, S.B., Schmitz, R.J. & Amasino, R.M. (2006). A PHD finger protein involved in both the vernalization and photoperiod pathways in Arabidopsis. *Genes & Development*, 20(23), pp. 3244-3248.

- Sutter, D. & Doerfler, W. (1980). Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. *Proc Natl Acad Sci U S A*, 77(1), pp. 253-6.
- Sutter, D., Westphal, M. & Doerfler, W. (1978). Patterns of integration of viral dna sequences in the genomes of adenovirus type 12-transformed hamster cells. *Cell*, 14(3), pp. 569-585.
- Tabara, M., Ohtani, M., Kanekatsu, M., Moriyama, H. & Fukuhara, T. (2018). Size Distribution of Small Interfering RNAs in Various Organs at Different Developmental Stages is Primarily Determined by the Dicing Activity of Dicer-Like Proteins in Plants. *Plant Cell Physiol*, 59(11), pp. 2228-2238.
- Takami, Y., Ono, T., Fukagawa, T., Shibahara, K. & Nakayama, T. (2007). Essential role of chromatin assembly factor-1-mediated rapid nucleosome assembly for DNA replication and cell division in vertebrate cells. *Mol Biol Cell*, 18(1), pp. 129-41.
- Talbert, P.B. & Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol*, 11(4), pp. 264-75.
- Tamada, Y., Yun, J.-Y., Woo, S.c. & Amasino, R.M. (2009). ARABIDOPSIS TRITHORAX-RELATED7 Is Required for Methylation of Lysine 4 of Histone H3 and for Transcriptional Activation of FLOWERING LOCUS C. *The Plant Cell*, 21(10), p. 3257.
- Tan, S. & Davey, C.A. (2011). Nucleosome structural studies. *Curr Opin Struct Biol*, 21(1), pp. 128-36.
- Tang, K., Lang, Z.B., Zhang, H. & Zhu, J.K. (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nature Plants*, 2(11).
- Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y. & Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. *Proc Natl Acad Sci U S A*, 100(15), pp. 8823-7.
- Thorstensen, T., Grini, P.E. & Aalen, R.B. (2011). SET domain proteins in plant development. *Biochim Biophys Acta*, 1809(8), pp. 407-20.
- Thorstensen, T., Grini, P.E., Mercy, I.S., Alm, V., Erdal, S., Aasland, R. & Aalen, R.B. (2008). The Arabidopsis SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS). *Plant Mol Biol*, 66(1-2), pp. 47-59.
- To, T.K., Kim, J.M., Matsui, A., Kurihara, Y., Morosawa, T., Ishida, J., Tanaka, M., Endo, T., Kakutani, T., Toyoda, T., Kimura, H., Yokoyama, S., Shinozaki, K. & Seki, M. (2011). Arabidopsis HDA6 Regulates Locus-Directed Heterochromatin Silencing in Cooperation with MET1. *PLoS Genetics*, 7(4).
- Trejo-Arellano, M.S., Mahrez, W., Nakamura, M., Moreno-Romero, J., Nanni, P., Köhler, C. & Hennig, L. (2017). H3K23me1 is an evolutionarily conserved histone modification associated with CG DNA methylation in Arabidopsis. *The Plant Journal*, 90(2), pp. 293--303.
- Trojer, P. & Reinberg, D. (2006). Histone lysine demethylases and their impact on epigenetics. *Cell*, 125(2), pp. 213-7.
- Tropberger, P. & Schneider, R. (2010). Going global Novel histone modifications in the globular domain of H3. *Epigenetics*, 5(2), pp. 112-117.

- Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G. & Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet*, 3(6), p. e86.
- Tuteja, N., Tran, N.Q., Dang, H.Q. & Tuteja, R. (2011). Plant MCM proteins: role in DNA replication and beyond. *Plant Mol Biol*, 77(6), pp. 537-45.
- van Attikum, H. & Gasser, S.M. (2009). Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol*, 19(5), pp. 207-17.
- van Nocker, S. & Ludwig, P. (2003). The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics*, 4(1), p. 50.
- Vardimon, L., Neumann, R., Kuhlmann, I., Sutter, D. & Doerfler, W. (1980). DNA methylation and viral gene expression in adenovirus-transformed and -infected cells. *Nucleic Acids Res*, 8(11), pp. 2461-73.
- Veiseth, S.V., Rahman, M.A., Yap, K.L., Fischer, A., Egge-Jacobsen, W., Reuter, G., Zhou, M.M., Aalen, R.B. & Thorstensen, T. (2011). The SUVR4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis. *PLoS Genet*, 7(3), p. e1001325.
- Veluchamy, A., Jegu, T., Ariel, F., Latrasse, D., Mariappan, K.G., Kim, S.K., Crespi, M., Hirt, H., Bergounioux, C., Raynaud, C. & Benhamed, M. (2016). LHP1 Regulates H3K27me3 Spreading and Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. *PLoS One*, 11(7), p. e0158936.
- Vilkaitis, G., Merkiene, E., Serva, S., Weinhold, E. & Klimasauskas, S. (2001). The mechanism of DNA cytosine-5 methylation. Kinetic and mutational dissection of HhaI methyltransferase. *J Biol Chem*, 276(24), pp. 20924-34.
- Waddington, C.H. (1957). *The strategy of the genes: A discussion of some aspects of theoretical biology*. London: Allen & Unwin.
- Wang, D., Tyson, M.D., Jackson, S.S. & Yadegari, R. (2006). Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 103(35), p. 13244.
- Wang, Z., Cao, H., Chen, F. & Liu, Y. (2014). The roles of histone acetylation in seed performance and plant development. *Plant Physiology and Biochemistry*, 84, pp. 125-133.
- Weaver, L.M. & Amasino, R.M. (2001). Senescence is induced in individually darkened Arabidopsis leaves, but inhibited in whole darkened plants. *Plant Physiol*, 127(3), pp. 876-86.
- Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L. & Kohler, C. (2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet*, 6(10).
- Widom, J. (1998). Structure, dynamics, and function of chromatin in vitro. *Annu Rev Biophys Biomol Struct*, 27, pp. 285-327.
- Williams, B. & Gehring, M. (2017). Stable transgenerational epigenetic inheritance requires a DNA methylation-sensing circuit. *Nature Communications*, 8.
- Williams, B.P., Pignatta, D., Henikoff, S. & Gehring, M. (2015). Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet*, 11(3), p. e1005142.

- Woo, H.R., Pontes, O., Pikaard, C.S. & Richards, E.J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization, 21(3), pp. 267-277.
- Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S. & Helliwell, C.A. (2006). The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc Natl Acad Sci U S A*, 103(39), pp. 14631-6.
- Wu, L., Mao, L. & Qi, Y.J. (2012). Roles of DICER-LIKE and ARGONAUTE Proteins in TAS-Derived Small Interfering RNA-Triggered DNA Methylation. *Plant Physiology*, 160(2), pp. 990-999.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. & Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*, 2(5), p. E104.
- Xu, L. & Shen, W.H. (2008). Polycomb Silencing of KNOX Genes Confines Shoot Stem Cell Niches in *Arabidopsis*. *Current Biology*, 18(24), pp. 1966-1971.
- Xu, L., Zhao, Z., Dong, A., Soubigou-Taconnat, L., Renou, J.P., Steinmetz, A. & Shen, W.H. (2008). Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. *Mol Cell Biol*, 28(4), pp. 1348-60.
- Yang, D.L., Zhang, G.P., Tang, K., Li, J.W., Yang, L., Huang, H., Zhang, H. & Zhu, J.K. (2016). Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. *Cell Research*, 26(1), pp. 66-82.
- Yang, H., Han, Z., Cao, Y., Fan, D., Li, H., Mo, H., Feng, Y., Liu, L., Wang, Z., Yue, Y., Cui, S., Chen, S., Chai, J. & Ma, L. (2012a). A Companion Cell-Dominant and Developmentally Regulated H3K4 Demethylase Controls Flowering Time in *Arabidopsis* via the Repression of FLC Expression. *PLoS Genetics*, 8(4), p. e1002664.
- Yang, H., Mo, H., Fan, D., Cao, Y., Cui, S. & Ma, L. (2012b). Overexpression of a histone H3K4 demethylase, JMJ15, accelerates flowering time in *Arabidopsis*. *Plant Cell Reports*, 31(7), pp. 1297-1308.
- Yang, J., Lior-Hoffmann, L., Wang, S., Zhang, Y. & Broyde, S. (2013). DNA cytosine methylation: structural and thermodynamic characterization of the epigenetic marking mechanism. *Biochemistry*, 52(16), pp. 2828-38.
- Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A. & Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*, 26(4), pp. 577-90.
- Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., Sung, Z.R. & Takahashi, S. (2001). EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in *Arabidopsis*. *Plant Cell*, 13(11), pp. 2471-81.
- Yoshinaga, S.K., Peterson, C.L., Herskowitz, I. & Yamamoto, K.R. (1992). Roles of Swi1, Swi2, and Swi3 Proteins for Transcriptional Enhancement by Steroid-Receptors. *Science*, 258(5088), pp. 1598-1604.
- Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L. & Zilberman, D. (2013). The *Arabidopsis* Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell*, 153(1), pp. 193-205.

- Zemach, A. & Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Current Biology*, 20(17), pp. R780-5.
- Zhang, H., Lang, Z. & Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, 19(8), pp. 489-506.
- Zhang, H.M., Tang, K., Qian, W.Q., Duan, C.G., Wang, B.S., Zhang, H., Wang, P.C., Zhu, X.H., Lang, Z.B., Yang, Y. & Zhu, J.K. (2014). An Rrp6-like Protein Positively Regulates Noncoding RNA Levels and DNA Methylation in Arabidopsis. *Molecular Cell*, 54(3), pp. 418-430.
- Zhang, K., Sridhar, V.V., Zhu, J., Kapoor, A. & Zhu, J.-K. (2007a). Distinctive core histone post-translational modification patterns in Arabidopsis thaliana. *PLoS One*, 2(11), pp. e1210-e1210.
- Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M. & Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol*, 10(6), p. R62.
- Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J. & Jacobsen, S.E. (2007b). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biol*, 5(5), p. e129.
- Zhang, X., Germann, S., Blus, B.J., Khorasanizadeh, S., Gaudin, V. & Jacobsen, S.E. (2007c). The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat Struct Mol Biol*, 14(9), pp. 869-71.
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E. & Ecker, J.R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*, 126(6), pp. 1189-201.
- Zheng, B. & Chen, X. (2011). Dynamics of histone H3 lysine 27 trimethylation in plant development. *Curr Opin Plant Biol*, 14(2), pp. 123-9.
- Zhong, X.H., Du, J.M., Hale, C.J., Gallego-Bartolome, J., Feng, S.H., Vashisht, A.A., Chory, J., Wohlschlegel, J.A., Patel, D.J. & Jacobsen, S.E. (2014). Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases. *Cell*, 157(5), pp. 1050-1060.
- Zhou, M., Palanca, A.M.S. & Law, J.A. (2018). Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nature Genetics*, 50(6), pp. 865-+.
- Zhou, S., Liu, X., Zhou, C., Zhou, Q., Zhao, Y., Li, G. & Zhou, D.X. (2016). Cooperation between the H3K27me3 Chromatin Mark and Non-CG Methylation in Epigenetic Regulation. *Plant Physiol*, 172(2), pp. 1131-1141.
- Zhou, Y., Tergemina, E., Cui, H., Forde, A., Hartwig, B., Velikkakam James, G., Schneeberger, K. & Turck, F. (2017). Ctf4-related protein recruits LHP1-PRC2 to maintain H3K27me3 levels in dividing cells in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 114(18), pp. 4833-4838.
- Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet*, 43, pp. 143-66.
- Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol*, 18(1), p. 87.
- Zilberman, D., Cao, X.F. & Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, 299(5607), pp. 716-719.
- Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature*, 456(7218), pp. 125-9.

Popular science summary

The instructions controlling the developmental transitions and responses to external stimuli in a living organism are written in the genome. The tangible representation of the genome is in the form of a DNA molecule. If the DNA would be extended, the linear string would be several meters long. How come this string can be fitted into the reduced space of a cell whose diameter is even smaller than that of a hair? This is accomplished by compacting the DNA through chemical associations with scaffold proteins that can expose or occlude the DNA, depending on the organism's demand to read a given instruction from the genome. The association between the DNA and those scaffold proteins gives rise to the chromatin. Additionally, the DNA itself can carry chemical modifications, like DNA methylation, that modulate multiple developmental transitions throughout the lifetime.

Inspired by the phenomenal display of colours observed in the aging leaves of plants during the autumn, we used the model plant *Arabidopsis thaliana* to investigate the changes in the chromatin occurring when the leaves become old. We corroborated that this transition is accompanied by partial expansion of the chromatin and few changes in DNA methylation. This reconfiguration is another example of the crucial role of the chromatin in preserving the genome instructions and controlling when they are going to be read. The scaffold proteins from the chromatin can also be chemically modified. We contributed with the discovery of two previously unknown modifications in the plant chromatin. By reading the distribution of the new modifications throughout the genome, we were able to show that their presence is associated with uncovering or occluding the DNA.

Together, this thesis contributed to a better understanding of the composition and modulation of the *Arabidopsis* chromatin and the resulting genome readout.

Resumen de divulgación

El conjunto de instrucciones que controlan el desarrollo y respuesta a estímulos externos en todo organismo vivo están escritas en su genoma. La representación tangible del genoma es la molécula de ADN. Si extendiéramos el ADN, resultaría en una larga cadena de varios metros en longitud. ¿Cómo entonces es que el ADN cabe en un espacio tan reducido como lo son las células, cuyo diámetro es incluso más estrecho que el de un cabello? Esto es posible gracias a que la asociación química entre el ADN y ciertas proteínas andamio resulta en la compactación del ADN. Dichas proteínas andamio pueden a su vez exponer u ocultar el ADN y, por lo tanto, exponer u ocultar la lista de instrucciones a llevar a cabo dependiendo de las necesidades del organismo. La asociación entre el ADN y las proteínas andamio conforma lo que se conoce como cromatina. Además, el ADN puede ser modificado químicamente, metilado por ejemplo, para modular múltiples transiciones del desarrollo. Inspirados por el espectáculo de cambio de color en las hojas de las plantas durante el otoño, utilizamos la planta modelo *Arabidopsis thaliana* para investigar los cambios que ocurren en la cromatina cuando las hojas envejecen. Corroboramos que dicha transición está acompañada por la expansión parcial de la cromatina, así como por pocos cambios en la metilación del ADN. Dicha reconfiguración representa un ejemplo más del rol crucial de la cromatina como centinela de las instrucciones del genoma. Las proteínas andamio de la cromatina también pueden estar sujetas a modificaciones químicas. En el presente estudio reportamos por primera vez la existencia de dos modificaciones en las proteínas andamio de plantas, y mostramos que su presencia está asociada con la exposición u oclusión del ADN. En conjunto, esta tesis contribuye al entendimiento sobre la composición y regulación de la cromatina en *Arabidopsis*, así como la correspondiente expresión del genoma.

Acknowledgements

First and foremost, I would like to acknowledge and thank my supervisor Prof. Dr. Lars Hennig, for giving me the chance to grow as a scientist and as person, guided by his committed mentorship through the hills and valleys of my PhD. To Prof. Dr. Claudia Köhler for her support towards the culmination of the journey and for inspiring me by example: with her passion and dedication towards science.

None of the discoveries presented in this thesis would have been possible without the neat work done by my collaborators in the wet lab. Thank you for your trust and patience.

Working at the Plant Biology Department has been a smooth experience thanks to the efficient work done by the janitors, technical and administrative staff. Thank you for your hard work.

"The bunny who lived on the moon, first arrived on a shooting star. When he looked down to check out the view, he realized he'd travelled quite far." – My special appreciation to the colleagues and friends, inside and outside the lab, who supported me and made me feel at home, no matter our relative latitudes.

The loving care I have received from my family has been vital to make the best out of every decision that has led me here. Thank you for always standing by my side.

Unpublished manuscript.

Transgenerational phenotype aggravation in CAF-1 mutants reveals parent-of-origin specific epigenetic inheritance

Iva Mozgova^{1,2*} , Thomas Wildhaber^{3*}, Minerva S. Trejo-Arellano¹, Jiri Fajkus⁴, Pawel Roszak³, Claudia Köhler¹ and Lars Hennig¹ 

¹Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, SE-75007 Uppsala, Sweden; ²Institute of Microbiology of the Czech Academy of Sciences, Centre Algatech, Opatovický mlýn, CZ-37981 Třeboň, Czech Republic; ³Department of Biology and Zurich-Basel Plant Science Center, ETH Zurich, CH-8092 Zurich, Switzerland; ⁴Laboratory of Functional Genomics and Proteomics, National Centre for Biomolecular Research, Faculty of Science, Masaryk University, CZ-61137 Brno, Czech Republic

Summary

Author for correspondence:

Lars Hennig

Tel: +46 18 673326

Email: Lars.Hennig@slu.se

Received: 20 December 2017

Accepted: 5 February 2018

New Phytologist (2018)

doi: 10.1111/nph.15082

Key words: *Arabidopsis thaliana*, CAF-1, Chromatin, Development, DNA methylation, epigenetics, histone.

- Chromatin is assembled by histone chaperones such as chromatin assembly factor CAF-1. We had noticed that vigor of *Arabidopsis thaliana* CAF-1 mutants decreased over several generations. Because changes in mutant phenotype severity over generations are unusual, we asked how repeated selfing of *Arabidopsis* CAF-1 mutants affects phenotype severity.
- CAF-1 mutant plants of various generations were grown, and developmental phenotypes, transcriptomes and DNA cytosine-methylation profiles were compared quantitatively.
- Shoot- and root-related growth phenotypes were progressively more affected in successive generations of CAF-1 mutants. Early and late generations of the *fasciata* (*fas*)2-4 CAF-1 mutant displayed only limited changes in gene expression, of which increasing upregulation of plant defense-related genes reflects the transgenerational phenotype aggravation. Likewise, global DNA methylation in the sequence context CHG but not CG or CHH (where H = A, T or C) changed over generations in *fas*2-4. Crossing early and late generation *fas*2-4 plants established that the maternal contribution to the phenotype severity exceeds the paternal contribution.
- Together, epigenetic rather than genetic mechanisms underlie the progressive developmental phenotype aggravation in the *Arabidopsis* CAF-1 mutants and preferred maternal transmission reveals a more efficient reprogramming of epigenetic information in the male than the female germline.

Introduction

Nuclear DNA is packaged into chromatin, which affects and regulates major cellular processes such as transcription, replication, DNA-repair and silencing of transposable elements (TEs). The basic unit of chromatin is the nucleosome, a hetero-octamer of histones H2A, H2B, H3 and H4, which organizes 147 bp of DNA. Chromatin assembly factor 1 (CAF-1) is a histone chaperone that initiates nucleosome formation by depositing H3–H4 dimers on free DNA after replication or DNA repair (Ramirez-Parra & Gutierrez, 2007b; Yu *et al.*, 2015). CAF-1 consists of three subunits, which are conserved in all eukaryotes (Verreault *et al.*, 1996; Kaufman *et al.*, 1997) and are called FASCIATA 1 (FAS1), FASCIATA 2 (FAS2) and MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) in *Arabidopsis* (Kaya *et al.*, 2001). Lack of CAF-1 is lethal in mammalian cells and causes developmental

arrest in *Drosophila* (Nabatiyan & Krude, 2004; Song *et al.*, 2007).

In *Arabidopsis*, *fas1* and *fas2* mutants are viable (Kaya *et al.*, 2001), whereas *msi1* mutants are embryo-lethal due to the crucial function of MSI1 in Polycomb repressive complexes (PRC2) (Köhler *et al.*, 2003; Guittion *et al.*, 2004). The *fas1* and *fas2* CAF-1 mutants develop several phenotypic defects, including stem fasciation, abnormal leaf and flower morphology, and disorganization of the shoot and root apical meristems (Reinholz, 1966; Leyser & Furner, 1992; Kaya *et al.*, 2001). CAF-1 mutants have also defects in cell fate specification (Costa & Shaw, 2006; Exner *et al.*, 2006). Recently, it has been shown that some of the developmental phenotypes are a result of unrestricted activation of defense genes (Mozgova *et al.*, 2015). Additionally, CAF-1 is required for the organization of heterochromatin and maintenance of transcriptional gene silencing, including inactivation of certain TEs (Kaya *et al.*, 2001; Kirik *et al.*, 2006; Ono *et al.*, 2006; Schönrock *et al.*, 2006), regulation of endoreduplication (Exner *et al.*, 2006; Kirik *et al.*, 2006; Ramirez-Parra &

*These authors contributed equally to this work.

Gutierrez, 2007a,b), regulation of cell cycle duration (Ramirez-Parra & Gutierrez, 2007a,b; Abe *et al.*, 2008; Chen *et al.*, 2008) and homologous recombination (Endo *et al.*, 2006; Kirik *et al.*, 2006). More recently, a functional link was found between CAF-1 and maintenance of telomeres and ribosomal DNA (rDNA) (Mozgova *et al.*, 2010). In *fas1* and *fas2*, telomere shortening and loss of rDNA occurs in an increasingly severe fashion over several generations. It also had been noticed that vigor of *fas1* and *fas2* mutants decreased over several generations (Mozgova *et al.*, 2010) but this effect was not thoroughly documented. As most mutants in genetic model systems have stable phenotypes, a possible change in phenotype severity over generations of the *fas1* and *fas2* CAF-1 mutants is remarkable. Here, we asked how repeated selfing of *fas1* and *fas2* affects phenotype severity. We found that plant size, juvenile–adult phase transition and maternal reproduction are progressively more affected in successive generations of the *fas1* and *fas2* CAF-1 mutants. Fully expanded leaves of early and late *fas2* generations display only limited changes in gene expression, of which upregulation of plant defense-related genes reflects the transgenerational phenotype aggravation. The developmental phenotypes of the CAF-1 mutants but not the tandem repeat copy-number are readily complemented by the presence of a functional CAF-1 complex. By crossing early and late generation mutants, we find that the maternal contribution to the phenotype severity exceeds the paternal contribution. Together, we establish that epigenetic rather than genetic mechanisms underlie the progressive developmental phenotype aggravation in the Arabidopsis CAF-1 mutants.

Materials and Methods

Plant material

All experiments used *Arabidopsis thaliana* (L.) Heynh. accession Col-0 and the *fas1-4* (SALK_N828822) and *fas2-4* (SALK_N533228) mutants (Exner *et al.*, 2006). For further information and scheme of plant propagation see Supporting Information Methods and Fig. S1. For the *fas1-4 35S::FAS1* lines, the *FAS1* cDNA was amplified (for primers see Table S1) and recombined into pMDC32 (Curtis & Grossniklaus, 2003) followed by transformation into *fas1* generation 4 (*G*₄) plants. The relative quantification of 45S rDNA repeat number and telomere length analysis was performed as described (Mozgova *et al.*, 2010).

Plants were grown at 21°C with humidity of 60% under 16 h : 8 h, or 8 h : 16 h, light : dark cycles corresponding to long-day or short-day conditions, respectively, at 120 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Plants under AGRONOMICS conditions (Baerenfaller *et al.*, 2012) were cultivated at 8 h 22°C : 16 h 21°C, light : dark short-day cycles, with humidity of 60% and a light intensity of 100 $\mu\text{mol m}^{-2} \text{s}^{-1}$. To expose plants to waterlogging stress, pots with 10-d-old plants were submerged in water up to soil level for the entire duration of the plant life. The submerging water was exchanged two times per week. Stressed plants were cultivated under long-day cycles as described above.

Rosette diameter, silique length, shoot length and length of silique number 6 were measured after plants were fully grown and carried only mature siliques. The juvenile–adult phase transition was determined by emerging of trichomes on the abaxial (lower) surface of rosette leaves. For details on characterization of ovule development see Methods S1.

Gene expression analysis

Reverse transcription polymerase chain reaction (RT-PCR) was performed as described previously (Mozgova *et al.*, 2015) (for primers see Table S1). For transcriptome analysis, wild-type (WT), *fas2-4 G*₁ and *G*₆ plants were grown for 48 d according to AGRONOMICS standard conditions (Baerenfaller *et al.*, 2012). Leaf number 6 was harvested at ZT (zeitgeber time) = 7 and processed for transcriptome profiling on Affymetrix AGRONOMICS1 Arabidopsis tiling arrays (Rehauer *et al.*, 2010) as described (Mozgova *et al.*, 2015). The experiment was performed in biological triplicates with each triplicate consisting of pooled leaves from five, seven and nine plants for WT Col, *fas2-4 G*₁ and *fas2-4 G*₆, respectively. Data were normalized and analyzed as described (Rehauer *et al.*, 2010; Müller *et al.*, 2012). Differential expression was tested with LIMMA (Smyth, 2004) followed by multiple testing correction (Storey & Tibshirani, 2003). Genes were considered to be differentially expressed when $q < 0.05$ and fold-change > 1.5 . Differentially expressed genes (DEGs) were assigned to one of eight principal classes of change by regression analysis. Briefly, gene expression vectors per gene were (0, 1) normalized before squared distances to each of the eight principal expression profiles were calculated. Each of the selected DEGs was assigned to the profile class for which the sum of squared distances was minimal.

Analysis of cytosine-DNA methylation by bisulfite sequencing

Bisulfite-converted DNA was sequenced in paired-end mode of 90 nucleotides long reads using an Illumina HiSeq2000 sequencer (Illumina Inc., San Diego, CA, USA). Quality control was performed with in-house scripts. Adapter trimming and removal of low complexity sequences was done using reaper (Davis *et al.*, 2013). Reads with total sequencing quality of < 30 Phred or shorter than 10 nucleotides were discarded. Clean reads were mapped to the reference genome TAIR10 using bismark (Krueger & Andrews, 2011) allowing at most one mismatch per 25-nucleotide seed. Forward and reverse reads were mapped independently (Table S2). Conversion rates and methylation status of cytosines were obtained using bismark_methylation_extractor. Cytosines in the plastid genome are not methylated, allowing estimates of bisulfite conversion efficiency. The mean conversion rate was 96% and the estimated mean false positive methylation rates were $< 5\%$ (Table S3). Cytosines that were covered by at least seven reads in WT, *G*₁ and *G*₆ were considered for the calculation of differentially methylated regions (DMRs). The Arabidopsis TAIR10 genome sequence was divided into windows of 100 bp; windows containing less than three covered

cytosines were discarded. To avoid calling DMRs with artificially small or large methylation differences due to low coverages, the 10th percentile of bins with lowest coverage was selected and bins with the 20% smallest and 20% largest average methylation values were discarded. Bins with a P -value ≤ 0.01 (Fisher's exact test) were considered as significant. Methylation changes are expressed as the relative difference between the given genotypes. DMRs were classified in eight nonoverlapping classes according to the transgenerational trends of change observed for DEGs.

Accession numbers

Data are available at GEO and ArrayExpress (accession numbers GSE104456 and E-MTAB-6136).

Results

Developmental phenotypes of CAF-1 mutants change over generations

In order to test for a potential change in phenotype severity in *Arabidopsis* Chromatin assembly factor 1 (CAF-1) mutants, *fas2-4* plants of the first to the sixth generation were grown in parallel (Fig. S1). Note that generation 1 (G_1) refers to the first homozygous mutant plants that segregate from a selfed heterozygous parental plant. There was a clear transgenerational aggravation of mutant plant vigor (Fig. 1a–c). The phenotype aggravation was visible as early as 13 d after induction of germination, as the plantlets of later generations appeared considerably smaller than those of earlier generations (Fig. 1a). This difference was even more pronounced at 17 (Fig. 1b) and 31 d (Fig. 1c). Reduced rosette and silique size and accelerated bolting were most obvious. Decreasing seed set led to hardly any seeds being produced by G_6 plants under our conditions.

Quantification of *fas2-4* developmental phenotypes showed a gradual transgenerational phenotype aggravation (Fig. 1d–h). Rosette diameter of fully grown plants decreased from c. 60% of WT rosette size in *fas2-4* G_1 to c. 25% in G_6 plants (Fig. 1d). The number of total rosette leaves decreased gradually from c. 90% of WT numbers in *fas2-4* G_1 to <60% in G_6 plants (Fig. 1e). Rosette leaf numbers in *fas2-4* were mainly affected by the number of adult leaves (Fig. 1f,g), which gradually decreased from c. 60% in *fas2-4* G_1 to <10% in G_6 plants (Fig. 1g). Frequently, G_6 plants completely lacked adult leaves and formed only juvenile leaves before bolting (Fig. 1f,g). These results showed a gradual reduction of total leaf number as a consequence of the gradual reduction of adult leaf numbers in later generations of *fas2-4* mutant plants. Similar effects of generation on rosette diameter and bolting were observed for *fas1-4* (see below, Fig. 6a), demonstrating that the transgenerational aggravation is neither specific for a particular allele nor for loss of FAS2 but is characteristic for disrupted CAF-1 function in *Arabidopsis*.

The *fas1* and *fas2* mutants have defects not only in above-ground organs originating from shoot apical meristems (SAM), but also in root growth and development (Leyser & Furner, 1992; Kaya *et al.*, 2001; Costa & Shaw, 2006; Ramirez-Parra &

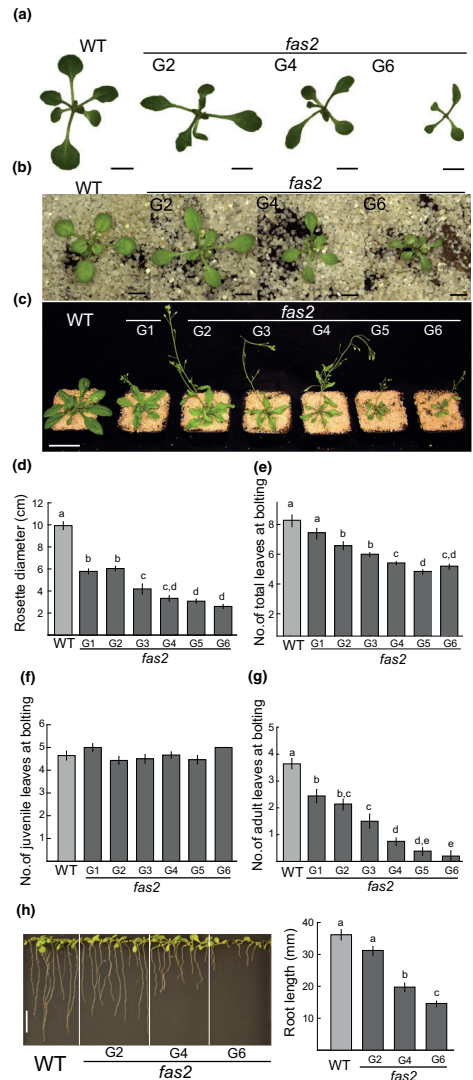


Fig. 1 Transgenerational phenotype aggravation in *Arabidopsis thaliana* *fas2* mutant plants. (a–c) Examples of phenotypes of plants (a) 13 d after germination (dag), (b) 17 dag and (c) 31 dag. (d) Quantification of rosette diameter of fully grown plants ($n = 14$ plants). (e–g) Quantification of the number of (e) total, (f) juvenile and (g) adult rosette leaves at bolting ($n = 14$ plants). (h) Example and quantification of root length in seedlings at 10 dag ($n = 25$ –30 seedlings grown on four different plates). Bars represent means \pm SE. Different letters above bars indicate significant difference ($P < 0.05$ in Student's t -test). Bars: (a, b) 5 mm; (c) 5 cm; (h) 1 cm. WT, wild-type; G, generation.

Gutierrez, 2007a,b; Pavlistova *et al.*, 2016). Therefore, we asked whether transgenerational aggravation also occurs for root phenotypes. Indeed, the impaired root growth of *fas2-4* became more severe in later generations (Fig. 1h). Together, both SAM-related and root apical meristem-related phenotypes show transgenerational aggravation in *fas2-4* plants, most likely due to disturbed CAF-1 function.

Defects in ovule development cause reduced seed set in *fas2-4*

As we had noticed reduced seed set in *fas2-4* plants, we next studied reproductive development in more detail. Because reduced seed set often causes reduced silique length, we measured silique length in several generations of *fas2-4* (Fig. 2a). Silique length greatly decreased from 70% in *fas2-4* G₁ to < 20% in G₆, demonstrating a transgenerational aggravation similar to that of other tested phenotypes. In order to identify reasons for reduced fruit length, seed set was quantified in naturally pollinated WT, *fas2-4* G₂ and G₆ plants (Fig. 2b). Under our experimental conditions, WT siliques had *c.* 45 normally developed seeds and five nonfertilized ovules. By contrast, siliques of *fas2-4* G₂ had *c.* 20 normally developed seeds and 10 unfertilized ovules. In *fas2-4* G₆, no normally developed seeds could be found and siliques contained 25 unfertilized ovules. Seed abortion was always below 1%. Thus, reduced silique length in *fas2-4* is caused by two effects. First, *fas2-4* forms *c.* 40% fewer ovules per flower than WT. This phenotype is almost completely penetrant in G₂ as it does not differ considerably between G₂ and G₆ (Fig. 2b). Second, lack of FAS2 caused reduced efficiency of fertilization. This phenotype is only partially visible in G₂ (30% of unfertilized ovules) but fully penetrant in G₆ (100% of unfertilized ovules). Together, reduced number of ovules formed and a large fraction of unfertilized ovules explain the short siliques and low seed set in *fas2-4* mutants.

Next, we investigated ovule development in more detail. Stages of WT ovule development include integument outgrowth, integument fusion and closure (Fig. 2c-i-ii), and finally formation of mature ovules (Fig. 2c-iii). In *fas2-4* G₂ plants (Fig. 2c-iv-viii), integuments of most ovules started to grow (Fig. 2c-iv), fused and closed normally (Fig. 2c-v). In some ovules, however, integuments did not fuse properly (Fig. 2c-vi) and tissue was protruding out of the ovule with the mature ovule sometimes bearing incompletely fused integuments (Fig. 2c-vii, viii). In G₆ plants (Fig. 2c-ix-xi), integuments grew only very incipiently (Fig. 2c-ix) before arresting prematurely (Fig. 2c-x). The unfused integuments determine defective ovules, which cannot be fertilized (Fig. 2c-xi). Note that no normally developed ovules could be found in G₆, consistent with the complete lack of developing seeds in these plants.

In summary, these results demonstrate that ovule development is strongly impaired in subsequent *fas2-4* generations, most likely leading eventually to complete sterility.

Transgenerational aggravation of transcriptome changes in *fas2-4* plants

In order to test whether the transgenerational aggravation of developmental phenotypes of *fas2-4* is also reflected in the

transcriptomes, we profiled gene expression in fully expanded rosette leaves of WT, *fas2-4* G₁ and G₆ mutant plants. Although there were only 62 significantly upregulated and 26 downregulated genes in *fas2-4* G₁ compared to WT, 295 and 62 genes were up- and downregulated, respectively, in G₆ (Fig. 3a; Table S4). Of the 62 genes upregulated in G₁, about one third also were found upregulated in G₆. Likewise, five of the 26 genes downregulated in G₁ also were downregulated in G₆. No genes were found to be dysregulated in an opposite manner between G₁ and G₆ (i.e. downregulated in G₁ and upregulated in G₆, or vice versa). Together, considerably more genes changed expression in the late than in the early generation of *fas2-4*. In total, 417 genes were affected in at least one mutant sample.

Although CAF-1 is needed for normal heterochromatin formation (Kirik *et al.*, 2006), previous studies did not find widespread activation of TE genes in Arabidopsis CAF-1 mutants (Schönrock *et al.*, 2006). The tiling array used here probes 2424 TE genes. Of these, only one (*AT4G04410*, a copia-like retrotransposon) was mildly upregulated in *fas2-4* G₁ and none in *fas2-4* G₆, consistent with maintained repression of heterochromatically silenced TE genes, even in late generations of *fas2-4*. Activation of *CACTA* TEs was observed in some cells in a fraction of mutant plants (Ono *et al.*, 2006). Because microarrays may lack the sensitivity to detect weak activation, we used RT-PCR to probe possible activation of *CACTA* TEs in seedlings of *fas2-4* G₂ and G₆. However, no transcripts could be detected for *CACTA* or the *TA2* TE in any mutant material (Fig. 3b). By contrast, transcripts of another silenced sequence (*TS1*) were readily detected in all *fas2-4* samples, consistent with earlier reports (Takeda *et al.*, 2004; Schönrock *et al.*, 2006). Note that *TS1* is not probed by the used microarray. The lack of strong signals for TEs *CACTA* and *TA2* is consistent with the reported rare stochastic activation (Ono *et al.*, 2006). Together, neither transcriptome nor RT-PCR data indicate an increasing loss of heterochromatic gene silencing in late generations of *fas2-4*.

Arabidopsis CAF-1 interacts with the repressive machinery of Polycomb group (PcG) proteins (Jiang & Berger, 2017). At a genome-wide scale, however, histone 3 lysine 27 trimethylation (H3K27me3)-positive PcG target genes were not preferentially enriched among genes upregulated in leaves of *fas2-4* G₁ or G₆ mutant plants (Fig. S2a). By contrast, H3K27me3-positive genes were considerably enriched among genes upregulated in *fas2* seedlings similar to the overlap between H3K27me3-positive genes and genes upregulated in H3.1-deficient plants. Because seedlings but not expanded leaves contain replicating cells, it is tempting to speculate that a CAF-1-PcG protein interaction enforces repression via H3K27me3 in dividing cells, but that eventually full repression by PcG proteins can be established even in the absence of CAF-1.

CAF-1 is considered to be the major histone chaperone for H3.1-H4, and both CAF-1 and H3.1 deficiency affect expression of similar numbers of genes (Jiang & Berger, 2017; Fig. 3a). There is, however, only limited overlap between genes upregulated in CAF-1 or H3.1-deficient plants (Fig. S2b). Although this overlap is statistically significant, the limited scale of overlap suggests that lack of CAF-1 affects Arabidopsis not only by reduced H3.1 incorporation but also by other means.

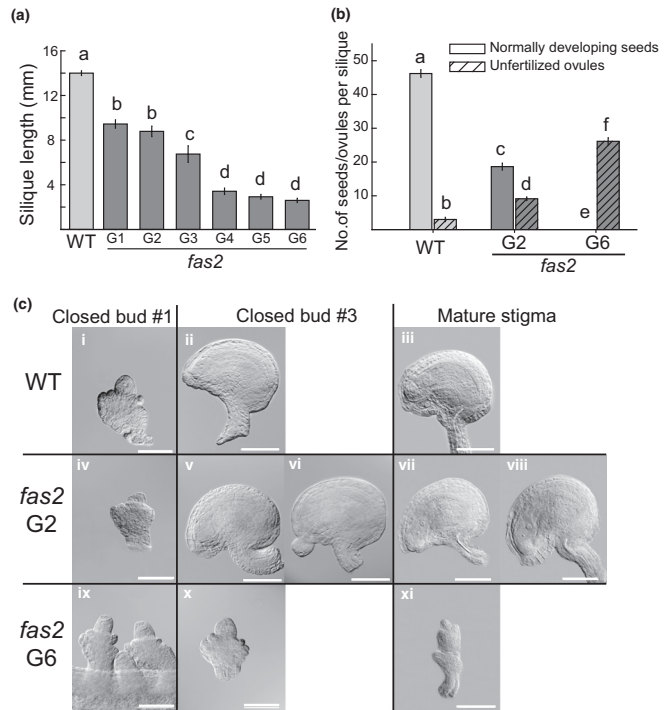


Fig. 2 Transgenerational loss of fertility in *Arabidopsis thaliana* *fas2* mutant plants. (a) Progressive reduction of the length of silique number 4 over six successive generations of *fas2*. Silique size was scored when all siliques on the plant were fully mature. (b) Reduced number of developing seeds and increasing number of unfertilized ovules in generations (G) G₂ and G₆ of *fas2*. Bars represent means \pm SE. Different letters above bars indicate significant difference ($P < 0.05$ in Student's *t*-test). (c) Transgenerational aggravation of ovule developmental phenotypes in *fas2*. Ovules were taken from gynoecea from the first and third youngest closed buds (closed bud #1, closed bud #3, resp.) and from gynoeceum with mature stigma. Bars, 50 μ m. WT, wild-type.

As we were interested in the pattern of changes across generations for the 417 genes with altered transcription, we assigned each gene to one of eight principal classes of change (see the Materials and Methods section for details; Table 1; Fig. 3c,d). A majority of the 330 upregulated genes (Fig. 3c, classes 1–4), belonged to class 2 – that is, was not affected in G₁ but mainly in G₆ (Tables 1, S5; Fig. 3c). Another considerable group of 84 upregulated genes was changed to some degree in G₁ but much more in G₆ (class 1). By contrast, only 46 genes were affected in a similar way in G₁ and G₆ (class 3). Finally, < 10% (26 genes) of the upregulated genes were affected mainly in G₁ but barely in G₆ (class 4). For the 87 downregulated genes (Fig. 3d, classes 5–8), the majority of 46 genes belonged to class 5 – that is, was affected in G₁ but even more in G₆. Only eight genes were affected in G₆ but not in G₁ (class 6). Eleven genes were affected in a similar way in G₁ and G₆ (class 7). Finally, 22 genes were downregulated in G₁ but not in G₆ (class 8). Together, most genes showed either a gradual increase in expression changes from G₁ to G₆ or were exclusively affected in G₆, consistent with the much more severe developmental phenotype of *fas2-4* G₆ than G₁ plants.

Gene ontology (GO) analysis revealed that upregulated genes of the dominating classes 1 ($\text{Expr}_{\text{Col}} < \text{Expr}_{\text{G1}} < \text{Expr}_{\text{G6}}$) and 2

($\text{Expr}_{\text{Col}} \approx \text{Expr}_{\text{G1}} < \text{Expr}_{\text{G6}}$) were significantly enriched for terms related to plant defense (Table 2). The most common group of downregulated genes (class 5, $\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G1}} > \text{Expr}_{\text{G6}}$) was enriched for the term ‘chloroplast’ and downregulated genes of class 8 ($\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G1}} < \text{Expr}_{\text{G6}}$) were enriched for terms related to abiotic stress. Activation of plant defense genes in *fas2-4* has been reported before (Mozgova *et al.*, 2015) and the present analysis shows that this activation is considerably more pronounced in *fas2-4* G₆ than in *fas2-4* G₁, revealing a transgenerational aggravation of distorted gene expression in the *fas2-4* CAF-1 mutant.

Together, activation of defense-related genes reflected the transgenerational aggravation of developmental phenotypes of *fas2-4*, whereas TE silencing appeared independent of the tested mutant generation.

Sequence context-specific transgenerational changes of DNA methylation in *fas2* mutant plants

Arabidopsis CAF-1 mutants interact synthetically with reduction in DNA cytosine methylation (Schönrock *et al.*, 2006) and DNA methylation was reported to be altered in CAF-1 mutants (Pontvianne *et al.*, 2013; Stroud *et al.*, 2013). Therefore, we asked

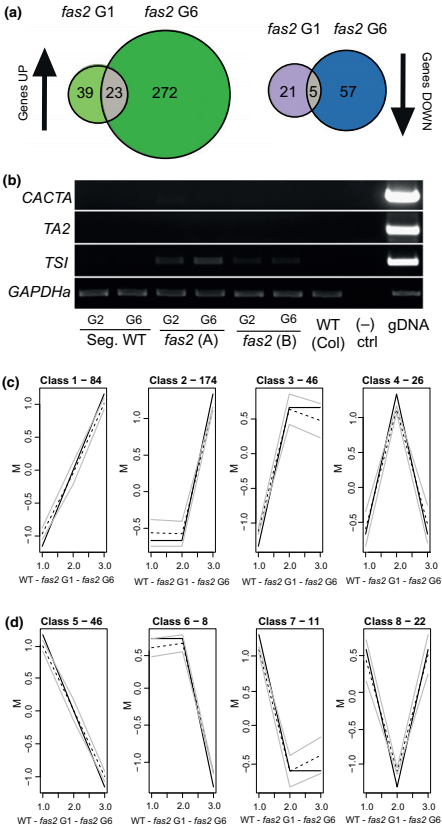


Fig. 3 Gene expression changes in increasing generations of *Arabidopsis thaliana* *fas2* mutant plants. (a) Venn diagram depicting numbers of genes up- and downregulated in fully expanded rosette leaves of *fas2* G₂ and *fas2* G₆ compared to wild-type (WT). (b) Transposable element expression in wild-type and *fas2* plants assayed by semi-quantitative reverse transcription polymerase chain reaction (RT-PCR). WT and mutant plants (*fas2*) were segregated from one *fas2*-4/+ heterozygous parent. WT (Col-0) plants (WT (Col)) served as control. PCR controls were no template control (–) ctrl and Col-0 genomic DNA (gDNA). (c) Classes of gene expression changes identified among the genes upregulated in *fas2* G₂ or *fas2* G₆ plants compared to WT. (d) Classes of gene expression changes identified among the genes downregulated in *fas2* G₂ or *fas2* G₆ plants compared to WT. (c, d) y-axis represents normalized gene expression change, x-axis represents the analysed genotypes (1, WT; 2, *fas2* G₁; 3, *fas2* G₆). Solid black lines in graphs represent the theoretical model, dashed black lines shows means of experimental expression values for all genes assigned to a particular model and gray lines represent 95% confidence intervals. G, generation.

whether DNA methylation undergoes a similar transgenerational change in *fas2* as developmental phenotypes and gene expression. Bisulfite sequencing of WT, and early (G₁) and late (G₆)

Table 1 Classification of differentially expressed genes according to expression patterns

Class	Description	Genes
1	$\text{Expr}_{\text{Col}} < \text{Expr}_{\text{G}_1} < \text{Expr}_{\text{G}_6}$	84
2	$\text{Expr}_{\text{Col}} \approx \text{Expr}_{\text{G}_1} < \text{Expr}_{\text{G}_6}$	174
3	$\text{Expr}_{\text{Col}} < \text{Expr}_{\text{G}_1} \approx \text{Expr}_{\text{G}_6}$	46
4	$\text{Expr}_{\text{Col}} < \text{Expr}_{\text{G}_1} > \text{Expr}_{\text{G}_6}$	26
5	$\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G}_1} > \text{Expr}_{\text{G}_6}$	46
6	$\text{Expr}_{\text{Col}} \approx \text{Expr}_{\text{G}_1} > \text{Expr}_{\text{G}_6}$	8
7	$\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G}_1} \approx \text{Expr}_{\text{G}_6}$	11
8	$\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G}_1} < \text{Expr}_{\text{G}_6}$	22

Genes found to be differentially expressed in *Arabidopsis thaliana* *fas2* generation (G₁ or *fas2* G₆ plants were assigned to one of eight main predefined classes (see the Materials and Methods section for details).

generations of the *fas2*-4 mutant was performed. Plotting the proportion of methylated cytosines (meCs) across all protein coding genes, no global differences in methylation were observed for any of the three sequence contexts CG, CHG and CHH (where H = A, T or C) (Fig. 4a). This established that gene body DNA methylation is not widely affected even in *fas2* G₆. By contrast, averaged DNA methylation on TEs was affected in *fas2*. Global CG DNA methylation on TEs did not differ between WT and *fas2* G₁, but was increased in *fas2* G₆ (Fig. 4a). Global CHG DNA methylation on TEs was higher in *fas2* G₁ than in WT and again higher in G₆ than in G₁ (Fig. 4a). Finally, CHH methylation was similar to or slightly lower than in WT in *fas2* G₁ and slightly higher than in WT in G₆ (Fig. 4a). Because CHH contexts are less frequent in the genome than CG or CHG contexts, the CHH methylation profiles are more variable.

Although methylation profiles reveal global trends that can be subtle in amplitude, differentially methylated region (DMRs) reveal larger changes that are locally restricted. We identified 15 522, 2335 and 8703 individual 50-bp DMRs for the CG, CHG and CHH contexts, respectively. Grouping the DMRs in similar classes as the DEGs, revealed that aggravation of differential DNA methylation changes was rare in all sequence contexts (classes 1 and 5) (Table S6). Most CG DMRs were changed in G₁ without major additional changes in G₆ (classes 3 and 7). Most CHG DMRs had increased methylation in G₁ without major additional changes in G₆ (class 3). CHH DMRs were likewise abundant in all classes without continuous change (i.e. all except for classes 1 and 5) (Table S6). Together, DMRs in *fas2* G₁ and G₆ do not generally reflect the global trends of methylation changes and rarely show transgenerational aggravation.

Although gene body DNA methylation was not globally altered, most CG DMRs come from gene bodies regardless of the class (Fig. 4b). By contrast, CHG and CHH DMRs often come from TEs. DMRs that map to TEs come most often from intergenic TEs and only rarely from TEs in promoters or gene bodies regardless of the methylation context (Fig. 4b). When TEs are grouped in families, CG DMRs of class 4 are enriched in the Long Terminal Repeat (LTR)-Copia, LTR-Gypsy and Rolling Circle (RC)-Helitron TE families (Fig. 4c). CHG DMRs of classes 1, 2, 4, 6 and 8 also are enriched in the LTR-Copia or

Table 2 Enrichment of gene ontology terms among differentially expressed genes in the eight defined classes

Class	Pattern	GO term	Obs. frequency	Exp. frequency	Enrichment	$-\log_{10}$ (P-value)
1	$\text{Expr}_{\text{Col}} < \text{Expr}_{\text{G1}} < \text{Expr}_{\text{G6}}$	Defence response to bacterium	6	0.5	3.7	4.1
2	$\text{Expr}_{\text{Col}} \approx \text{Expr}_{\text{G1}} < \text{Expr}_{\text{G6}}$	Systemic acquired resistance	6	0.1	5.21	7.9
3	$\text{Expr}_{\text{Col}} < \text{Expr}_{\text{G1}} \approx \text{Expr}_{\text{G6}}$	Plant-type cell wall	4	0.4	3.35	2.3
4	$\text{Expr}_{\text{Col}} < \text{Expr}_{\text{G1}} > \text{Expr}_{\text{G6}}$	Anchored to membrane	4	0.2	4.31	4
5	$\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G1}} > \text{Expr}_{\text{G6}}$	Chloroplast	20	4.1	2.28	8.1
6	$\text{Expr}_{\text{Col}} \approx \text{Expr}_{\text{G1}} > \text{Expr}_{\text{G6}}$	No significant GO terms				
7	$\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G1}} \approx \text{Expr}_{\text{G6}}$	No significant GO terms				
8	$\text{Expr}_{\text{Col}} > \text{Expr}_{\text{G1}} > \text{Expr}_{\text{G6}}$	Cold acclimation	3	0	7.69	7.2
		Response to water deprivation	5	0.1	5.4	7
		Response to cold	5	0.2	4.95	6.2

Shown are observed frequency, expected frequency, enrichment and $-\log_{10}$ (P-value) for all gene ontology (GO) terms significantly over-represented ($P < 0.01$) among the genes of a class with differential expression in in *Arabidopsis thaliana* *fas2* G₁ or *fas2* G₆ plants.

LTR-Gypsy families. CHH DMRs of most classes show the same tendency to be enriched in LTR-Gypsy family but they are also often enriched in RC-Helitrons (Fig. 4c). Together, DMRs in *fas2* affect various TE families but are particularly enriched in LTR TEs. Because LTR TEs are typical for pericentromeric heterochromatin, lack of CAF-1 seems to affect DNA methylation most in pericentromeric heterochromatin. This is consistent with the reported role of CAF-1 in organization of heterochromatin (Kirik *et al.*, 2006) but it does not seem to be reflected by global transcriptional activation of pericentromeric TEs.

Abiotic stress increases the severity of some *fas2-4* phenotypes

Activation of defense-related genes, which are responsive to biotic stress, has been shown to increase the severity of the CAF-1 mutant phenotypes (Mozgova *et al.*, 2015). Here, we asked whether abiotic stress can have a similar effect, contributing to the variable severity of the *fasciata* phenotypes. Wild-type, *fas2-4* G₂ and *fas2-4* G₄ were cultivated under control conditions or subjected to waterlogging stress – an abiotic stress treatment that could be well controlled even for soil-grown plants. This stress treatment affected plant growth and development and reduced plant vigor (Fig. 5a,b) in all tested genetic backgrounds. Stressed *fas2-4* plants showed an additional aggravation of the phenotype concerning number of adult leaves, rosette diameter and silique length (Fig. 5b). In detail, stress reduced the number of adult leaves from 65% and 40% of WT numbers in control plants of *fas2-4* G₂ and G₄, respectively, to 20% and 0% in stressed plants. Likewise, silique length was reduced from 70% and 45% in control *fas2-4* plants to 35% and 20% in stressed plants. Although the waterlogging stress severely affected rosette size in both WT and *fas2-4* (reducing the rosette size by 70–80% relative to the respective nonstressed control in both genotypes and generations), it had a less severe effect on the number of adult leaves and the silique length in WT (reduced by 40% and 20%, respectively) than in *fas2-4* (reduced by 60% and 70–75%, respectively), suggesting that the latter phenotypic traits are more vulnerable to be affected specifically in the *fas2-4* background

than rosette size. These results established that the specific quantitative *fas2-4* phenotype strongly depends on growth conditions, that is, rosette or silique size in *fas2-4* can differ between experiments but the pattern of transgenerational aggravation is consistently observed. Together, stressed *fas2-4* plants of an earlier generation became similar to *fas2-4* plants of a later generation. In other words, waterlogging stress could mimic continuous selfing of *fas2-4* plants.

Next, we asked whether the stress-induced aggravation of the phenotype was heritable. Seeds were harvested from stressed and unstressed plants and progeny was cultivated under control condition. In these plants, rosette size and silique length were assessed (Fig. 5c). There were no significant differences between progeny from stressed and unstressed plants for any tested genotype. In particular, the severity of the *fas2-4* phenotype was determined only by the generation of selfing and not by previous exposure to the abiotic stress.

Alleles causing developmental phenotypes of CAF1 mutants are unstable in the presence of CAF1

The described experiments had established that loss of Arabidopsis CAF-1 leads to a transgenerational aggravation of the mutant phenotype. Next, we asked whether the aggravated phenotype can persist even in the presence of reintroduced CAF-1 function. We had noted that *fas1-4* CAF-1 mutant plants showed a similar transgenerational phenotype aggravation as *fas2-4* plants (Fig. 6a) and compared *fas1-4* G₂, G₄ and G₆ plants to *fas1-4* 35S::*FAS1* plants. Because the 35S::*FAS1* transgene was introduced into *fas1-4* G₄ plants and maintained for two generations to obtain T₂ plants for analysis, three main scenarios could be expected: (1) introduction of the *FAS1* transgene prevents further transgenerational phenotype aggravation; the analyzed *fas1-4* 35S::*FAS1* T₂ and *fas1-4* G₄ plants will be comparable. (2) Introduction of a *FAS1* transgene does not affect transgenerational phenotype aggravation; the analyzed *fas1-4* 35S::*FAS1* T₂ and *fas1-4* G₆ plants will be comparable. (3) Introduction of a *FAS1* transgene reverses transgenerational phenotype aggravation; the analyzed *fas1-4* 35S::*FAS1* T₂ plants will be similar to WT or plants of an

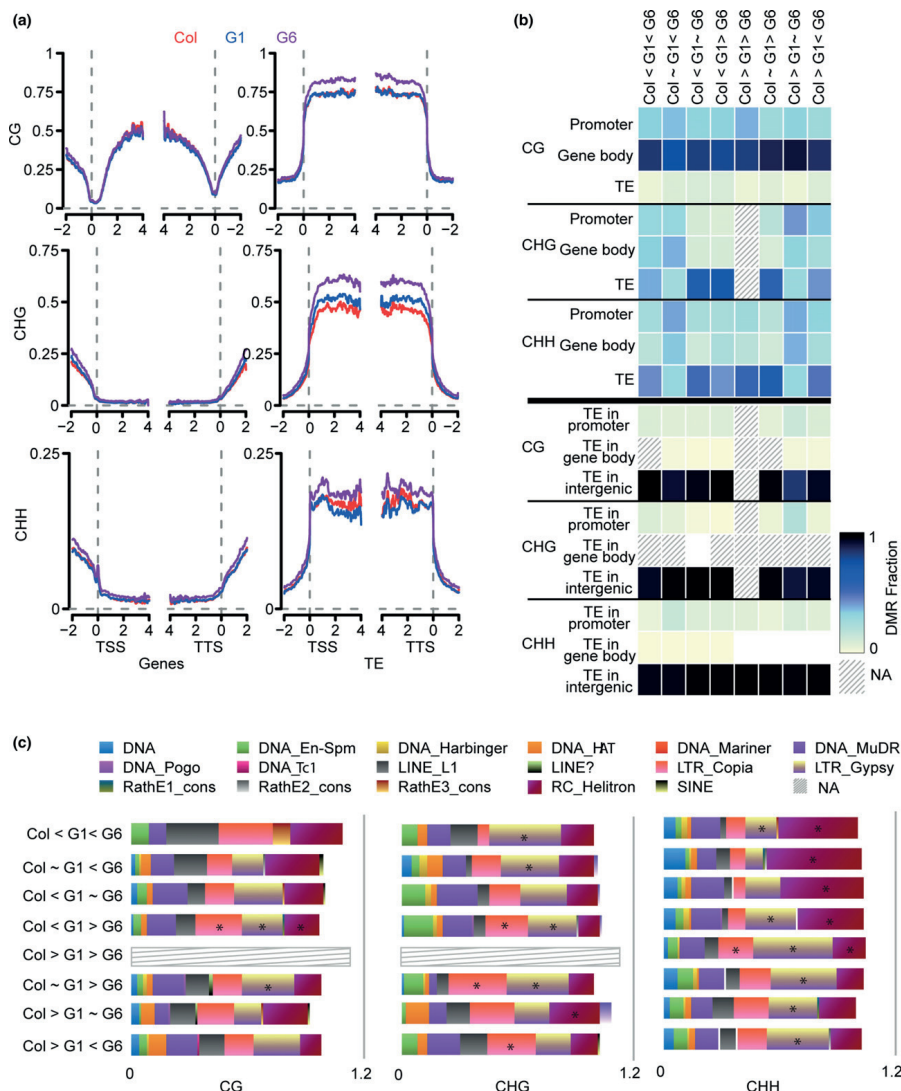


Fig. 4 Transgenerational changes of DNA methylation in *Arabidopsis thaliana* *fas2* mutant plants. (a) Metagene plots of raw methylation values along genes (left column) and transposable elements (TEs) (right column) across generations. Average values were plotted from 2 kb upstream and downstream of the transcriptional start site (TSS) and the transcriptional termination site (TTS), respectively, and 4 kb into the feature body for wild-type Col (red), *fas2* G₁ (blue) and *fas2* G₆ (purple) in the three methylation contexts. (b) Heatmap of the fraction of differentially methylated regions (DMRs) mapping to promoters, genes or different genomic locations of TEs. The fraction of DMRs from each class of change mapping to promoters (2 kb upstream of TSS), gene body or different genomic locations of TEs was calculated with respect to the total number of significant (Fisher's exact test, P -value ≤ 0.01) DMRs following the different trends of changes across generations. Underpopulated classes with ≤ 10 DMRs were not taken into account (shaded). The fraction of DMRs mapping to TEs was further classified into TE families. Different colors represent the fraction of DMRs per class mapping to a given TE family. TE families with a significant overpopulation (hypergeometric test, P -value ≤ 0.05) of DMRs are indicated with an asterisk (*). Underpopulated families with ≤ 10 DMRs were not taken into account (shaded). Col, wild-type; G, generation.

early *fas1-4* generation. Using two independent transgenic lines, the results clearly supported scenario (3). The *fas1-4* 35S::*FAS1* T₂ plants, although G₆ for *fas1-4*, had only a mild mutant phenotype that was even less severe than in *fas1-4* G₂ plants (Fig. 6a). Notably, the presence of the 35S::*FAS1* transgene did not complement the telomere length or the 45S rDNA repeat copy number in these transgenic lines (Fig. S3). Together, the aggravated phenotype of *fas1-4* plants after recurrent selfing is unstable in the presence of functional *FAS1*, despite the fact that the loss of tandem repeats in *fas1* is not fully reverted upon *FAS1* reintroduction (here and Pavlistova *et al.*, 2016). The reversibility of the phenotype indicates an epigenetic nature of the aggravated phenotype.

The aggravated phenotype in later generations could be caused by different (epi)allelic states at one or more loci. To test whether such allelic states could be characterized genetically, *fas2-4* G₁ and G₄ plants were crossed using the later generation as pollen donor (G₁ × G₄) (Figs S4, 6b). We asked whether the F₂ progeny of such a cross would segregate plants with phenotypes of distinct severity or resemble *fas2-4* G₃ or G₆ plants. However, no clear phenotypic classes of F₂ plants were evident and rosette diameters as well as silique lengths had similar spread (measured as standard deviation) for F₂ plants and regular *fas2-4* G₃ or G₆ plants. F₂ plants had significantly larger rosettes and siliques than *fas2-4* G₆ plants and significantly smaller rosettes and silique length than *fas2-4* G₃ plants. This argues against a single Mendelian locus underlying the analyzed developmental phenotype with transgenerational aggravation in *fas2-4*. For comparison, we also performed a cross of *fas2-4* G₁ and G₄ plants using the earlier generation as pollen donor (G₄ × G₁) (Fig. S4), and analyzed the F₂ progeny. Similar to the reciprocal G₁ × G₄ cross, again no clear phenotypic classes of F₂ plants were evident and rosette diameters as well as silique lengths had similar spread (measured as SD) for F₂ plants from the G₄ × G₁ cross and regular *fas2-4* G₃ or G₆ plants. In contrast to the reciprocal cross, however, average rosette diameters and silique lengths of G₄ × G₁ F₂ and *fas2-4* G₆ plants were similar and significantly smaller than that of *fas2-4* G₃ plants. Thus, the aggravated phenotype of G₄ plants is more efficiently transmitted through the maternal than through the paternal parent. Likewise, the phenotype of F₁ plants was more severely affected in a *fas2-4* G₄ × G₁ than in the reciprocal G₁ × G₄ cross (Fig. S5) supporting the notion of impaired transmission of the aggravated phenotype through the father.

Together, the dependency of the aggravated phenotype of Arabidopsis CAF-1 mutants on continuous lack of CAF-1 function and the unequal parental efficiency to transmit the aggravated phenotype indicate that epigenetic mechanisms underlie the observed transgenerational phenotypic aggravation.

Discussion

Here, we describe a parental-specific transgenerational aggravation of developmental and molecular phenotypes in Arabidopsis chromatin assembly factor 1 (CAF-1) mutants. The developmental phenotypes include reduced rosette size and root length, shortened adult phase and early flowering, reduced silique length,

and reduced ovule number. Furthermore, ovule development defects accumulated progressively, correlating with low fertilization efficiency and failure to produce seeds in late generations of the mutants. Molecular phenotypes with transgenerational aggravation were observed for CHG DNA methylation at transposable elements (TEs) and at the transcriptome level. We show that the CAF-1 mutant developmental phenotypes are influenced not only by the level of biotic stress (Mozgova *et al.*, 2015), but also by the abiotic growth conditions (here). Interestingly, ovule development defects in *fasciata 2* (*fas2*) closely resemble those caused by reduced amount of MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) in the *MSI1* co-suppression (*msi1-cs*) lines (Henig *et al.*, 2003). Therefore, we propose that undisturbed CAF-1 activity is needed for normal ovule development in Arabidopsis.

Developmental genes are not enriched among genes misregulated in seedlings or leaves of either early or late generation CAF-1 mutants (Schönrock *et al.*, 2006; Mozgova *et al.*, 2015; this work). Instead, we show here that stress-responsive genes are most enriched among genes that show progressive transgenerational upregulation in *fas2*. Stress-responsive genes also are enriched among genes affected by nucleosome depletion in *fas2* (Munoz-Viana *et al.*, 2017) and lack robust transcriptional repression in *fas1* and *fas2* mutant plants (Mozgova *et al.*, 2015). Stable repression of stress-responsive genes may thus be a general role of Arabidopsis CAF-1. The severity of developmental changes in *fas2* correlates with the amplitude of salicylic acid (SA) signaling, and reduced SA content can partially normalize *fas2* development (Mozgova *et al.*, 2015). In wild-type (WT) Arabidopsis, stress-induced chromatin changes are usually not heritable but rapidly reset (Probst & Mittelsten, 2015; Lamke & Baurle, 2017), whereas in some mutants resetting of stress-induced chromatin states is impaired (Iwasaki & Paszkowski, 2014). Because CAF-1 also is required for efficient resetting (Pecinka *et al.*, 2010), it appeared possible that intensified stress responses together with failure to reset stress-induced chromatin states underlie the transgenerational aggravation of the CAF-1 mutant phenotype. However, stress conditions for parental *fas2* plants did not affect the phenotype severity of the progeny and the additional phenotype aggravation in *fas2* that was induced by stress was not heritable. This suggests that the *fas2* mutant phenotype is determined by two components: a stress-related component that is not heritable and a stress-unrelated component that is heritable and shows transgenerational aggravation. This notion is consistent with earlier reports that impaired SA signaling could only mitigate but not fully suppress the *fas2* mutant phenotype (Mozgova *et al.*, 2015).

Transgenerational aggravation of developmental phenotypes in Arabidopsis mutants also has been reported for telomere maintenance mutants (Riha *et al.*, 2001). Progressive reduction of telomere length in mutants of the catalytic subunit of telomerase (TERT) is associated with phenotype aggravation, which has been hypothesized to be a consequence of increasing genome instability (Riha *et al.*, 2001). The CAF-1 dysfunction in Arabidopsis causes selective loss of tandem repetitive DNA sequences, including the 45S rDNA and the telomeres (Mozgova *et al.*, 2010), which could potentially contribute to the

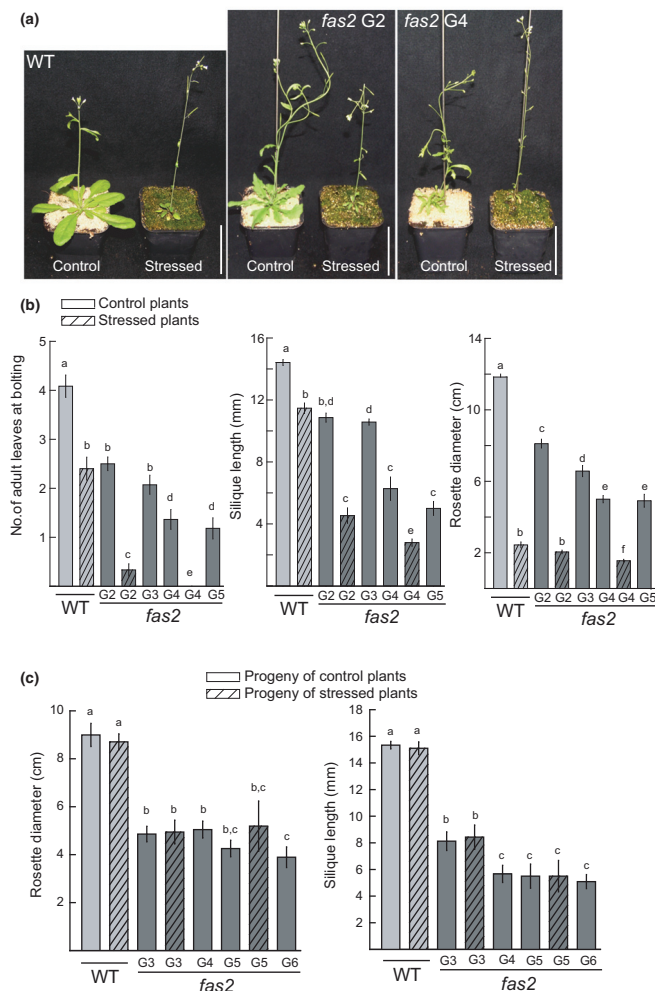


Fig. 5 Abiotic stress enhances the severity of *Arabidopsis thaliana* *fas2* mutant developmental phenotype but its effect is not heritable. (a) Examples of wild-type (WT) and *fas2-4* phenotypes of plants grown under control conditions (control) and exposed to waterlogging stress (stressed). Bars, 5 cm. (b) Quantification of phenotypes of control and waterlogged plants. (N_{control} = 14 plants, N_{stressed} = 21 plants). (c) Quantification of phenotypes of the progeny of control and waterlogged plants (N_{control} progeny = 14 plants, N_{stressed} progeny = 14 plants). Bars represent means \pm SE. Different letters above bars indicate significant difference ($P < 0.05$ in Student's *t*-test). G, generation.

progressive developmental phenotype severity in the CAF-1 mutants. However, as we show here, late generation *fas1* transformed with *35S::FAS1* display an early generation or even WT developmental phenotype, despite retention of low levels of tandem repetitive sequences in the genome. Neither reduced telomere length nor depletion of 45S rDNA is therefore likely to directly underlie the developmental phenotype or its aggravation in the CAF-1 mutants. The fast reversal of the *fas1* to the WT phenotype upon restoration of CAF-1 activity as also observed in (Pavlistova *et al.*, 2016) furthermore argues against genetic causes of the progressive developmental phenotype aggravation in the

CAF-1 mutants. Together with the parent-of-origin effect in the inheritance of alleles determining the early- or late-generation phenotypes, we propose that epigenetic rather than genetic defects underlie the developmental phenotype aggravation in the CAF-1 mutants.

How could lack of CAF-1 activity affect the epigenome? There are several well-established mechanisms that provide plausible scenarios. First, lack of CAF-1 causes locally reduced nucleosome occupancy or nucleosome gaps (Munoz-Viana *et al.*, 2017). This may lead to a more widespread loosening of chromatin packing. Second, lack of CAF-1 shifts the ratio of histone variants H3.1

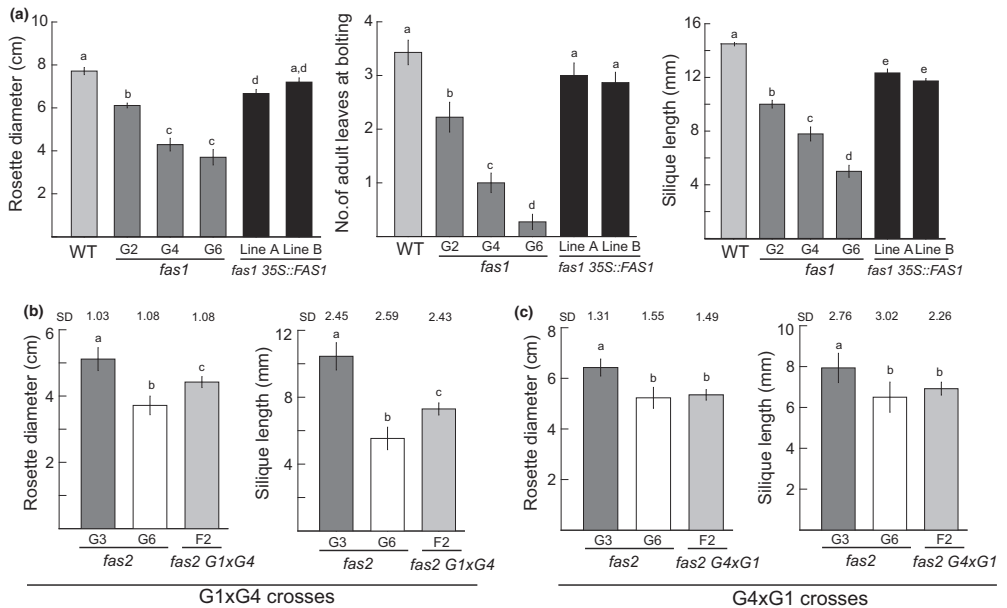


Fig. 6 Epigenetic mechanisms underlie the transgenerational phenotype aggravation in *Arabidopsis thaliana* chromatin assembly factor 1 (CAF-1) mutants. (a) Alleles causing developmental phenotypes of CAF-1 mutants are unstable in the presence of CAF-1. Quantification of phenotypes in two independent transgenic lines (line A, line B) of T₂ generation of *fas1-4 35S::FAS1* transformed in G₄ (i.e. equivalent of G₆ nontransformed plants *fas1-4*). (b) Phenotype quantification of the F₂ generation of plants originating from *fas2-4* G₁ × *fas2-4* G₄ crosses and control progeny of parental plants propagated in parallel (G₃ and G₆) (N_{F2} = 45, N_{G3,G6} = 15). (c) Phenotype quantification of the F₂ generation of plants originating from *fas2-4* G₄ × *fas2-4* G₁ crosses and control parental plants propagated in parallel (G₃ and G₆) (N_{F2} = 45, N_{G3,G6} = 15). Bars represent means ± SE. Different letters above bars indicate significant difference ($P < 0.05$ in a one-tailed Student's *t*-test). WT, wild-type; G, generation.

and H3.3 in chromatin, because only CAF-1-dependent chromatin assembly strongly prefers H3.1 over H3.3, whereas CAF-1-independent chromatin assembly, which can partially substitute for CAF-1 function in CAF-1 mutants, works well with H3.3 (Duc *et al.*, 2015, 2017). Third, CAF-1 locates to the site of the replication fork through its interaction with PROLIFERATING CELL NUCLEAR ANTIGEN (Shibahara & Stillman, 1999; Jiang & Berger, 2017) and lack of CAF-1 causes S-phase defects and replication stress, which impairs nucleosome-mediated epigenetic inheritance (Li *et al.*, 2017). Most histone modifications are re-established immediately after S-phase (Alabert *et al.*, 2015), often mediated by CAF-1 interacting with other chromatin proteins such as the mammalian methyl CpG binding protein that recruits a H3K9 methyltransferase (Sarraf & Stancheva, 2004) and HETEROCHROMATIN PROTEIN 1 (Quivy *et al.*, 2004). Plant CAF-1 interacts also with the repressive machinery of Polycomb group (PcG) proteins, thus contributing to maintenance of gene repression by PcG proteins (Jiang & Berger, 2017) at least in seedlings. It is likely that replication stress and impaired recruitment of chromatin proteins together underlie the requirement of CAF-1 for

the inheritance of epigenetically determined chromatin states (Monson *et al.*, 1997; Enomoto & Berman, 1998; Tchenio *et al.*, 2001; Ono *et al.*, 2006; Song *et al.*, 2007).

In *Arabidopsis*, where DNA cytosine methylation is abundant, each of the three described consequences of CAF-1 dysfunction on chromatin can affect DNA methylation. First, chromatin compaction by linker histone H1 and nucleosome density limit the access of DNA methyltransferases to their DNA substrate. This effect is largest for CHROMOMETHYLASE 3 (CMT3), which is responsible for CHG methylation (Zemach *et al.*, 2013). It is thus possible that increased CHG methylation in *fas2* is a consequence of reduced nucleosome occupancy in the absence of CAF-1. Although we did not observe a strong overlap between nucleosome depletion and changes in DNA methylation in the analyzed G₀ cells of *fas2* leaves, it is possible that transiently reduced nucleosome occupancy shortly after S-phase suffices to ease access of CMT3. Subsequent activity of CAF-1 independent nucleosome assembly mechanisms re-establishes wild-type-like nucleosome occupancy in most of the genome, as seen in the resting cells of leaves (Munoz-Viana *et al.*, 2017). Second, H3.3 limits H1 recruitment and thus favors DNA

methylation (Wollmann *et al.*, 2017). The shifted H3.1-H3.3 balance in CAF-1 mutants is thus expected to lead to reduced H1 presence and increased CHG methylation, which is consistent with our findings. Finally, replication stress and CAF-1 deficiency may reduce histone H3 lysine 9 dimethylation (H3K9me2) levels (Sarraf & Stancheva, 2004) and affect DNA methylation via the H3K9me2 feed-back loop.

We note that transgenerational phenotype aggravation can also occur when CG DNA methylation is reduced such as in mutants of DECREASED DNA METHYLATION 1 (DDM1), a chromatin remodeler required for DNA methylation (Kakutani *et al.*, 1996), and in mutants for the maintenance DNA methyltransferase MET1 (Mathieu *et al.*, 2007). In contrast to *fas2*, developmental defects in *ddm1* and *met1* are highly stochastic and can differ greatly between sibling plants. In *ddm1*, particular combinations of phenotypes were found more often than other combinations (Kakutani *et al.*, 1996). In *fas2*, the different aspects of the developmental phenotype consistently occurred together and the quantitative variability among plants of the same generation was generally low. In addition, *fas2* does not show the global reduction of CG methylation found in *ddm1* and *met1*. Because of the nonstochastic nature of the *fas2* phenotype and the lack of globally reduced CG methylation, we consider it unlikely that global CG DNA methylation changes similar to those that occur in *ddm1* or *met1* underlie the phenotype aggravation in CAF-1 mutants.

The inequality of reciprocal crosses between early G₁ and late G₄ generation *fas2* mutants suggests that the epigenetic determinants of phenotype severity are more efficiently transmitted through the egg than the sperm. Which properties make Arabidopsis sperm chromatin particular and could explain transmission differences? Current knowledge suggests two major scenarios:

First, sperm cell chromatin comprises mainly pollen-specific histone H3 variants, especially H3.10 (Borg & Berger, 2015). Although H3.3 and H3.1 differ only at four of 135 positions, H3.10 differs at 13 positions from H3.1. Thus, H3 protein properties such as stability in the nucleosome or efficiency to be targeted by histone modifiers is expected to vary considerably between pollen H3.10 and the sporophytic variants H3.1 and H3.3. In particular, Borg & Berger (2015) suggested that substitutions adjacent to K27 may impair K27 trimethylation by PRC2 for H3.10. Thus, epigenetic information contained in H3K27me3 may not be efficiently transmitted through sperm cells. Another characteristic feature of sperm cell chromatin is a loss of CHH methylation while CG and CHG methylation is maintained. CHH methylation is restored only after fertilization guided by 24-nt small interfering RNAs (Calarco *et al.*, 2012). Thus, epigenetic information contained in CHH methylation may be transmitted efficiently in sporophytic but not through sperm cells. Although only CHG and not CHH methylation showed global transgenerational changes in *fas2* plants, c. 8700 regions with locally altered CHH methylation were found that could, in principle, relate to the epigenetic transgenerational aggravation of the *fas2* phenotype. A recent report described that epigenetic memory of abiotic stress may be mediated by

epigenetically labile sites (Wibowo *et al.*, 2016). Similar to the determinants of *fas2* phenotype severity, these stress responses are transmitted much more efficiently maternally than paternally, an effect that was attributed to widespread DNA de-methylation in the male germline (Wibowo *et al.*, 2016). Future work will show whether DNA de-methylation in the male germline or the male-specific histone H3.10 forms a more efficient barrier to limit epigenetic inheritance through the paternal side in Arabidopsis. Finally, it is possible that preferred maternal determination of CAF-1 mutant phenotype severity as shown here is not directly related to chromatin properties but to other maternal effects such as steering seed development. However, although stress treatment greatly reduced maternal vigor, it failed to affect the offspring phenotype severity in our system, making it rather unlikely that reduced maternal vigor of *fas2* plants underlies the preferred maternal determination of CAF-1 mutant phenotype severity. In addition, the observed partial paternal transmission argues against an exclusive chromatin-independent maternal effect. Regardless of the molecular mechanism, the preferential maternal determination of Arabidopsis CAF-1 mutant phenotype severity strongly supports the notion that mothers have stronger non-genetic effects on offspring phenotypes than fathers. This is consistent with ecological scenarios that transgenerational phenotype plasticity can be adaptive when responding to maternal growth conditions (Galloway & Etterson, 2007).

Acknowledgements

This work was supported by grants from the Swiss National Science Foundation SNF, the Czech Science Foundation (16-01137S) and the Swedish Science Foundation VR and the Knut-and-Alice-Wallenberg Foundation.

Author contributions

L.H., I.M., T.W., J.F. and C.K. designed the research; T.W., I.M. and P.R. performed research; M.S.T.-A., I.M. and L.H. analyzed data; I.M. and L.H. wrote the manuscript; and I.M. and T.W. contributed equally this work.

ORCID

Iva Mozgova  <http://orcid.org/0000-0002-3815-9223>

Lars Hennig  <http://orcid.org/0000-0002-6645-1862>

References

- Abe M, Kuroshita H, Umeda M, Itoh J, Nagato Y. 2008. The rice flattened shoot meristem, encoding CAF-1 p150 subunit, is required for meristem maintenance by regulating the cell-cycle period. *Developmental Biology* 319: 384–393.
- Alabert C, Barth TK, Reveron-Gomez N, Sidoli S, Schmidt A, Jensen ON, Imhof A, Groth A. 2015. Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes & Development* 29: 585–590.
- Baerenfaller K, Massonnet C, Walsh S, Baginsky S, Buhlmann P, Hennig L, Hirsch-Hoffmann M, Howell KA, Kahlau S, Radziejewski A *et al.* 2012.

- Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. *Molecular Systems Biology* 8: 606.
- Borg M, Berger F. 2015. Chromatin remodelling during male gametophyte development. *Plant Journal* 83: 177–188.
- Calarco JP, Borges F, Donoghue MT, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijo JA, Becker JD *et al.* 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* 151: 194–205.
- Chen Z, Tan JL, Ingouff M, Sundaresan V, Berger F. 2008. Chromatin assembly factor 1 regulates the cell cycle but not cell fate during male gametogenesis in *Arabidopsis thaliana*. *Development* 135: 65–73.
- Costa S, Shaw P. 2006. Chromatin organization and cell fate switch respond to positional information in Arabidopsis. *Nature* 439: 493–496.
- Curtis MD, Grossniklaus U. 2003. A gateway cloning vector set for high-throughput functional analysis of genes in *planta*. *Plant Physiology* 133: 462–469.
- Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. 2013. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63: 41–49.
- Duc C, Benoit M, Detourne G, Simon L, Poulet A, Jung M, Veluchamy A, Latrasse D, Le Goff S, Cotterell S *et al.* 2017. Arabidopsis ATRX modulates H3.3 occupancy and fine-tunes gene expression. *Plant Cell* 29: 1773–1793.
- Duc C, Benoit M, Le Goff S, Simon L, Poulet A, Cotterell S, Tatout C, Probst AV. 2015. The histone chaperone complex HIR maintains nucleosome occupancy and counterbalances impaired histone deposition in CAF-1 complex mutants. *Plant Journal* 81: 707–722.
- Endo M, Ishikawa Y, Osakabe K, Nakayama S, Kaya H, Araki T, Shibahara KI, Abe K, Ichikawa H, Valentine L *et al.* 2006. Increased frequency of homologous recombination and T-DNA integration in Arabidopsis CAF-1 mutants. *EMBO Journal* 25: 5579–5590.
- Enomoto S, Berman J. 1998. Chromatin Assembly Factor 1 contributes to the maintenance, but not the re-establishment, of silencing at the yeast silent mating loci. *Genes & Development* 12: 219–232.
- Exner V, Taranto P, Schönrock N, Grussem W, Hennig L. 2006. Chromatin assembly factor CAF-1 is required for cellular differentiation during plant development. *Development* 133: 4163–4172.
- Galloway LF, Ettersson JR. 2007. Transgenerational plasticity is adaptive in the wild. *Science* 318: 1134–1136.
- Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, Berger F. 2004. Identification of new members of FERTILISATION INDEPENDENT SEED Polycomb group pathway involved in the control of seed development in *Arabidopsis thaliana*. *Development* 131: 2971–2981.
- Hennig L, Taranto P, Walser M, Schönrock N, Grussem W. 2003. Arabidopsis MS1 is required for epigenetic maintenance of reproductive development. *Development* 130: 2555–2565.
- Iwasaki M, Paszkowski J. 2014. Identification of genes preventing transgenerational transmission of stress-induced epigenetic states. *Proceedings of the National Academy of Sciences, USA* 111: 8547–8552.
- Jiang D, Berger F. 2017. DNA replication-coupled histone modification maintains Polycomb gene silencing in plants. *Science* 357: 1146–1149.
- Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ. 1996. Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proceedings of the National Academy of Sciences, USA* 93: 12 406–12 411.
- Kaufman PD, Kobayashi R, Stillman B. 1997. Ultraviolet radiation sensitivity and reduction of telomeric silencing in *Saccharomyces cerevisiae* cells lacking Chromatin Assembly Factor-1. *Genes & Development* 11: 345–357.
- Kaya H, Shibahara K, Taoka K, Iwabuchi M, Stillman B, Araki T. 2001. FASCIATA genes for Chromatin Assembly Factor-1 in Arabidopsis maintain the cellular organization of apical meristems. *Cell* 104: 131–142.
- Kirik A, Pecinka A, Wendler E, Reiss B. 2006. The Chromatin Assembly Factor subunit FASCIATA1 is involved in homologous recombination in plants. *Plant Cell* 18: 2431–2442.
- Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Grussem W. 2003. Arabidopsis MS1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO Journal* 22: 4804–4814.
- Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571–1572.
- Lamke J, Baule I. 2017. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology* 18: 124.
- Leyser HM, Furner IJ. 1992. Characterisation of three shoot apical meristem mutants of *Arabidopsis thaliana*. *Development* 116: 397–403.
- Li W, Yi J, Agbu P, Zhou Z, Kelley RL, Kallgren S, Jia S, He X. 2017. Replication stress affects the fidelity of nucleosome-mediated epigenetic inheritance. *PLoS Genetics* 13: e1006900.
- Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J. 2007. Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell* 130: 851–862.
- Monson EK, de Bruijn D, Zakian VA. 1997. The yeast CAC1 protein is required for the stable inheritance of transcriptionally repressed chromatin at telomeres. *Proceedings of the National Academy of Sciences, USA* 94: 13081–13086.
- Mozgova I, Mokros P, Fajkus J. 2010. Dysfunction of Chromatin Assembly Factor-1 induces shortening of telomeres and loss of 45S rDNA in *Arabidopsis thaliana*. *Plant Cell* 22: 2768–2780.
- Mozgova I, Wildhaber T, Liu Q, Abou-Mansour E, L'Haronid F, Métraux J, Grussem W, Hofius D, Hennig L. 2015. Chromatin assembly factor CAF-1 represses priming of plant defence response genes. *Nature Plants* 1: 15 127.
- Müller M, Patrignani A, Rehrauer H, Grussem W, Hennig L. 2012. Evaluation of alternative RNA labeling protocols for transcript profiling with Arabidopsis AGRONOMIC1 tiling arrays. *Plant Methods* 8: 18.
- Munoz-Viana R, Wildhaber T, Trejo-Arellano MS, Mozgova I, Hennig L. 2017. Arabidopsis Chromatin Assembly Factor 1 is required for occupancy and position of a subset of nucleosomes. *Plant Journal* 92: 363–374.
- Nabatiyan A, Krude T. 2004. Silencing of Chromatin Assembly Factor-1 in human cells leads to cell death and loss of chromatin assembly during DNA synthesis. *Molecular and Cellular Biology* 24: 2853–2862.
- Ono T, Kaya H, Takeda S, Abe M, Ogawa Y, Kato M, Kakutani T, Scheid OM, Araki T, Shibahara K. 2006. Chromatin Assembly Factor-1 ensures the stable maintenance of silent chromatin states in Arabidopsis. *Genes to Cells* 11: 153–162.
- Pavlistova V, Dvorackova M, Jez M, Mozgova I, Mokros P, Fajkus J. 2016. Phenotypic reversion in *Arabidopsis thaliana* by reintroduction of FAS genes: variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes. *Plant Journal* 88: 411–424.
- Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O. 2010. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell* 22: 3118–3129.
- Pontvianne F, Blevins T, Chandrasekhara C, Mozgova I, Hassel C, Pontes OM, Tucker S, Mokros P, Muchova V, Fajkus J *et al.* 2013. Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epialleles. *Genes & Development* 27: 1545–1550.
- Probst AV, Mittelsten Scheid O. 2015. Stress-induced structural changes in plant chromatin. *Current Opinion in Plant Biology* 27: 8–16.
- Quivy JP, Roche D, Kirschner D, Tagami H, Nakatani Y, Almouzni G. 2004. A CAF-1 dependent pool of HP1 during heterochromatin duplication. *EMBO Journal* 23: 3516–3526.
- Ramirez-Parra E, Gutierrez C. 2007a. E2F regulates FASCIATA1, a chromatin assembly gene whose loss switches on the endocycle and activates gene expression by changing the epigenetic status. *Plant Physiology* 144: 105–120.
- Ramirez-Parra E, Gutierrez C. 2007b. The many faces of Chromatin Assembly Factor 1. *Trends in Plant Science* 12: 570–576.
- Rehrauer H, Aquino C, Grussem W, Henz SR, Hilson P, Laubinger S, Naouar N, Patrignani A, Rombauts S, Shu H *et al.* 2010. AGRONOMIC1: a new resource for Arabidopsis transcriptome profiling. *Plant Physiology* 152: 487–499.
- Reinholz E. 1966. Radiation induced mutants showing changed inflorescence characteristics. *Arabidopsis Information Service* 3: 19–20.
- Riha K, McKnight TD, Griffing LR, Shippen DE. 2001. Living with genome instability: plant responses to telomere dysfunction. *Science* 291: 1797–1800.
- Sarrafa SA, Stancheva L. 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular Cell* 15: 595–605.

- Schönrock N, Exner V, Probst A, Gruissem W, Hennig L. 2006. Functional genomic analysis of CAF-1 mutants in *Arabidopsis thaliana*. *Journal of Biological Chemistry* 281: 9560–9568.
- Shibahara K, Stillman B. 1999. Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell* 96: 575–585.
- Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 3: 1–26.
- Song Y, He F, Xie G, Guo X, Xu Y, Chen Y, Liang X, Staglar I, Egli D, Ma J *et al.* 2007. CAF-1 is essential for Drosophila development and involved in the maintenance of epigenetic memory. *Developmental Biology* 311: 213–222.
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA* 100: 9440–9445.
- Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152: 352–364.
- Takeda S, Tadele Z, Hofmann I, Probst AV, Angelis KJ, Kaya H, Araki T, Mengiste T, Scheid OM, Shibahara K *et al.* 2004. BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis. *Genes & Development* 18: 782–793.
- Tchenio T, Casella JF, Heidmann T. 2001. A truncated form of the human CAF-1 p150 subunit impairs the maintenance of transcriptional gene silencing in mammalian cells. *Molecular and Cellular Biology* 21: 1953–1961.
- Verreault A, Kaufman PD, Kobayashi R, Stillman B. 1996. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* 87: 95–104.
- Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J *et al.* 2016. Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife* 5: e13546.
- Wollmann H, Stroud H, Yelagandula R, Tarutani Y, Jiang D, Jing L, Jamge B, Takeuchi H, Holec S, Nie X *et al.* 2017. The histone H3 variant H3.3 regulates gene body DNA methylation in *Arabidopsis thaliana*. *Genome Biology* 18: 94.
- Yu Z, Liu J, Deng WM, Jiao R. 2015. Histone chaperone CAF-1: essential roles in multi-cellular organism development. *Cellular and Molecular Life Sciences* 72: 327–337.
- Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153: 193–205.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Scheme of plant propagation and cultivation.

Fig. S2 CAF-1 deficiency does not cause widespread activation of H3K27me3-positive genes.

Fig. S3 Characterization of two independent T₂ *fas1 35S::FAS1* lines complemented in the fourth mutant generation (G4).

Fig. S4 Scheme of reciprocal crosses between early (G1) and late (G4) generation *fas2-4* mutants.

Fig. S5 Reciprocal crosses between early (G1) and late (G4) generation *fas2-4* result in different phenotype severity in populations of F₁ progeny.

Table S1 Primers used in this study

Table S2 Read statistics of the Bisulphite sequencing

Table S3 Estimation of the Bisulphite conversion rate

Table S4 List of up- and downregulated genes

Table S5 List of genes in defined classes

Table S6 Classification of differentially methylated regions expressed genes according to patterns of change

Methods S1 Supporting information on production of successive *fas1* and *fas2* mutant generations, characterization of ovule development and gene ontology analysis.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

H3K23me1 is an evolutionarily conserved histone modification associated with CG DNA methylation in *Arabidopsis*

Minerva S. Trejo-Arellano^{1,†} , Walid Mahrez^{1,2,†}, Miyuki Nakamura¹, Jordi Moreno-Romero¹, Paolo Nanni³, Claudia Köhler¹ and Lars Hennig^{1,*}

¹Department of Plant Biology and Linnean Center for Plant Biology, Swedish University of Agricultural Sciences, PO-Box 7080, Uppsala SE-75007, Sweden,

²Department of Biology and Zurich-Basel Plant Science Center, ETH Zurich, Zurich CH-8092, Switzerland, and

³Functional Genomics Center Zurich, University of Zurich/ETH Zurich, Zurich CH-8057, Switzerland

Received 16 September 2016; revised 9 December 2016; accepted 16 January 2017.

*For correspondence (email lars.hennig@slu.se).

†These authors contributed equally to this work.

SUMMARY

Amino-terminal tails of histones are targets for diverse post-translational modifications whose combinatorial action may constitute a code that will be read and interpreted by cellular proteins to define particular transcriptional states. Here, we describe monomethylation of histone H3 lysine 23 (H3K23me1) as a histone modification not previously described in plants. H3K23me1 is an evolutionarily conserved mark in diverse species of flowering plants. Chromatin immunoprecipitation followed by high-throughput sequencing in *Arabidopsis thaliana* showed that H3K23me1 was highly enriched in pericentromeric regions and depleted from chromosome arms. In transposable elements it co-localized with CG, CHG and CHH DNA methylation as well as with the heterochromatic histone mark H3K9me2. Transposable elements are often rich in H3K23me1 but different families vary in their enrichment: LTR-Gypsy elements are most enriched and RC/Helitron elements are least enriched. The histone methyltransferase KRYPTONITE and normal DNA methylation were required for normal levels of H3K23me1 on transposable elements. Immunostaining experiments confirmed the pericentromeric localization and also showed mild enrichment in less condensed regions. Accordingly, gene bodies of protein-coding genes had intermediate H3K23me1 levels, which coexisted with CG DNA methylation. Enrichment of H3K23me1 along gene bodies did not correlate with transcription levels. Together, this work establishes H3K23me1 as a so far undescribed component of the plant histone code.

Keywords: *Arabidopsis thaliana*, epigenetics, histone post-translational modifications, mass spectrometry, ChIP-seq, DNA methylation, heterochromatin, gene body methylation.

INTRODUCTION

Histones are small, highly basic proteins. Two pairs of H3, H4, H2A and H2B make up an octamer, around which about 147 bp of DNA are wrapped to build up the basic unit of chromatin – the nucleosome. Histones carry diverse covalent post-translational modifications (PTMs) such as acetylation, phosphorylation, methylation, ubiquitination and ADP-ribosylation (Bannister and Kouzarides, 2011). These dynamic and often evolutionarily conserved modifications act in combination to establish distinct chromatin states (Jenuwein and Allis, 2001). Histone PTMs can be read and interpreted by specific proteins, which

subsequently influence numerous biological processes including transcription, replication, chromosome maintenance and cell division.

Unlike histone lysine acetylation, which can affect DNA-protein interaction, histone lysine methylation is an epigenetic modification that does not alter the charge of the modified lysines on histone tails. Lysines can be mono- (me1), di- (me2) or trimethylated (me3) on their ϵ -amino group. Using mass spectrometry, methylation of the histone H3 lysines K4, K9, K27, K36, K79 and the H4 lysine K20 was found in plants (For review see Liu *et al.*, (2010a).

Depending on the target lysine and the degree of methylation, this histone PTM can be associated with active or inactive genes and can have different cytological localizations (Liu *et al.*, 2010a). Although these patterns are often conserved, the function or localization of histone PTMs can differ between different groups of eukaryotes. In mice, for example, H3K9me3 and H4K20me3 are most abundant in condensed heterochromatin and are associated with gene silencing whereas H3K9me1 and H3K9me2 are enriched in euchromatin (Loidl, 2004). In contrast, in *Arabidopsis* H3K9me3 and H4K20me3 are located in euchromatin and associated with gene expression (Charron *et al.*, 2009) while H3K9me1 and H3K9me2 are localized in heterochromatin and associated with gene silencing (Fransz *et al.*, 2006). Histone lysine methylation has important roles in many biological processes including control of transcription, cell-cycle regulation, DNA damage and stress responses, heterochromatin formation and X-chromosome inactivation (Martin and Zhang, 2005; Greer and Shi, 2012).

The homeostasis of the methylation on histone lysines is maintained by histone lysine methyl transferases (HKMT), which catalyze the transfer of methyl groups from S-adenosylmethionine to the ϵ -amino group of lysine, and by histone demethylases, which remove the methylation mark (Liu *et al.*, 2010a; Black *et al.*, 2012). Deposition of H3K9me2 in *Arabidopsis*, for instance, is catalyzed by the SET domain protein KRYPTONITE (KYP) (Jackson *et al.*, 2004). Once established, methylated lysines constitute landmarks for recognition by the aromatic cages of proteins with methyl-binding domains (Taverna *et al.*, 2007) such as PHD fingers, WD40 repeats, CW domains (Hoppmann *et al.*, 2011), PWWP domains, ankyrin repeats (Collins *et al.*, 2008), chromodomains, double chromodomains, chromobarrels, Tudor domains, double or tandem Tudor domains and MBT repeats. These protein readers often recruit additional protein complexes to perform functions such as gene silencing (Greer and Shi, 2012).

The addition of a methyl group to cytosine bases of the DNA to form 5-methylcytosine, referred to as DNA methylation, is another layer of epigenetic control that can stabilize repressed chromatin domains. DNA methylation occurs in both prokaryotes and eukaryotes. Although absent in budding and fission yeast as well as in *Caenorhabditis elegans*, this mark is prominent in fungi, plants and vertebrates (He *et al.*, 2011). In plants, DNA methylation occurs in the contexts of CG, CHG and CHH (H = A, C, or T) and is highly abundant in pericentromeric regions as well as other repetitive elements. DNA methylation in the CG sequence context also occurs in the transcribed region of nearly one-third of expressed *Arabidopsis* genes (Law and Jacobsen, 2010). Several studies have revealed the complex relationship between DNA methylation and histone PTMs. In *Arabidopsis*, loss of the DNA methyltransferase MET1 causes a significant

reduction of H3K9me2 at heterochromatic loci. Conversely, the reduction of the H3K9me2 levels in *kyp* mutants causes hypo-methylation on many transposable elements (Zhou, 2009). *In vitro* assays have revealed the binding of KYP to CG methylated DNA via its SRA (SET and RING associated) domain (Johnson *et al.*, 2007). Moreover, once H3K9me2 has been established by KYP, it can act as a binding site for the DNA methyltransferases CMT3 and CMT2, which methylate cytosines in CHG or CHH contexts, respectively (Law and Jacobsen, 2010). Furthermore, genome-wide profiling has revealed a strong correlation between DNA and H3K9 methylation (Bernatavichute *et al.*, 2008). In addition, it was found that DNA methylation coexists not only with H3K9me2 but also with H4K20me1, H3K27me1 and H3K27me2 (Roudier *et al.*, 2011). This suggests that DNA and histone lysine methylation cooperate in transcriptional gene regulation.

Despite detailed knowledge about many histone PTMs, new modifications are continuously being discovered, demonstrating the limitations of current knowledge (Chen *et al.*, 2007; Fujiki *et al.*, 2011; Tan *et al.*, 2011). Compared with yeast and animals, the histone PTMs in plants are even less well defined. For an in-depth understanding of chromatin function and for future modeling approaches, comprehensive lists of histone PTMs need to be established, at least for model organisms. In a previous study we used high-sensitivity mass spectrometry to search for additional PTMs on the amino-terminal tail of plant H3. We identified H3K36ac, which had not yet been characterized in plants (Mahrez *et al.*, 2016). Here, we describe H3K23me1 as a plant histone PTM. H3K23me1 is evolutionarily conserved among flowering plants. It is highly enriched in heterochromatin but is also found on some euchromatic coding genes. KYP is required for full accumulation of H3K23me1. In addition, H3K23me1 was found to depend on DNA methylation.

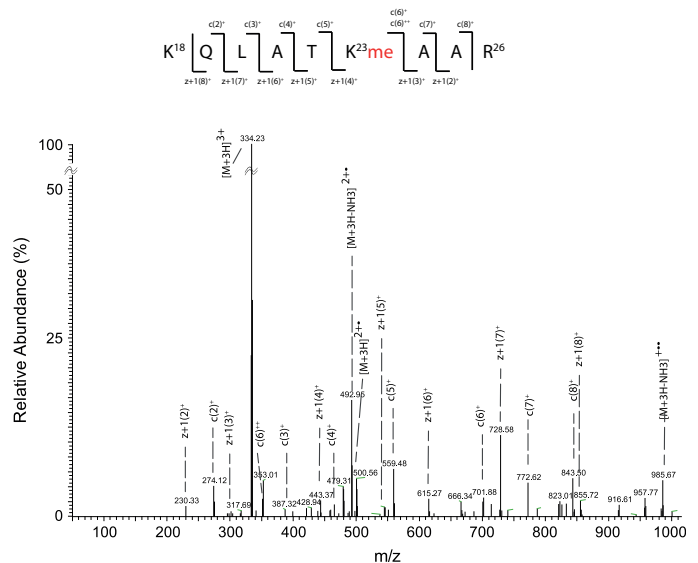
RESULTS

The histone modification H3K23me1 exists in plants

In a previous publication (Mahrez *et al.*, 2016) we reported the combination of high-throughput techniques to detect and characterize PTMs in H3 in plants, using *Brassica oleracea*. The manual validation of the results confirmed the identification of H3K23me1, which had not been previously described in plants. The tandem mass spectrometry (MS/MS) electron transfer dissociation spectrum of the ArgC-digested histone H3 in Figure 1 shows the *c* and *z* fragmentation ions obtained for the peptide KQLATK(me1)AAR (observed mass at *m/z* 334.2138) in its triply charged form, which allowed to unequivocally identify a monomethylation in lysine 23 of H3. Importantly, we were also able to detect the corresponding lysine 23 carrying an acetylation on other H3 peptides (Figure S1 in the Supporting

Figure 1. Identification of H3K23me1 from *Brassica oleracea*.

Tandem mass spectrometry fragmentation spectrum of the peptide QKLATK(me1)AAR (m/z 334.2183, 3⁺). This peptide was identified as the H3 peptide (histone H3, amino acids 18–26) derived from ArgC-digested RP-HPLC-purified cauliflower H3. Above the spectrum is the peptide sequence where the identified c-type ions and z-type ions, which contain the C terminus, are reported.



Information), confirming that this residue can be either methylated or acetylated. In a previous report using human NIH3T3 cells, H3K23 was found to be mono-, di- or trimethylated (Liu *et al.*, 2010b). Similarly, H3K23me1, H3K23me2 and H3K23me3 were found in *C. elegans* (Vandamme *et al.*, 2015). In our mass spectrometry study with plant histones we only detected the monomethylated form.

To verify the mass spectrometry results and confirm the existence of the histone modification, a commercial polyclonal anti-H3K23me1 antibody was used for immunoblotting assays with a histone extract from *Arabidopsis thaliana* inflorescences. This antibody was previously tested for specificity using peptide arrays as described in Liu *et al.* (2010b). A single band at about 17 kDa, corresponding to the size of H3, was obtained (Figure 2a). Together, these results confirm the existence of H3K23me1 in *Arabidopsis* and suggest that this modification is conserved among Brassicaceae.

H3K23me1 is enriched in heterochromatin

To investigate the localization of H3K23me1 in the nucleus, paraformaldehyde-fixed *Arabidopsis* root-tip nuclei were immunostained using the anti-H3K23me1 antibody. While some H3K23me1 fluorescence signal was present throughout the euchromatic regions, the most intense signal localized to the 4',6-diamidino-2-phenylindole (DAPI)-dense chromocenters which contain centromeric and pericentromeric heterochromatic sequences (Figure 2b–e). This observation was similar to the staining pattern observed for the heterochromatic histone mark H3K9me2 (Jackson

et al., 2004), and suggests that H3K23me1 is also mainly targeted to condensed and transcriptionally silent regions.

Genome-wide profiling of H3K23me1

To assess the distribution of the H3K23me1 enrichment along the genome in higher resolution, native chromatin immunoprecipitation (N-ChIP) followed by high-throughput sequencing was performed. Nucleosome density was normalized for by the anti-H3 signal. Plotting of the normalized H3K23me1 signal along chromosomes showed the highest values in the pericentromeric regions (Figure S2). Interestingly, higher levels of H3K23me1 (red line in Figure S2) correlated with a higher density of transposable elements (TEs) (orange line in Figure S2). Globally, H3K23me1 was also present along the chromosome arms but at much lower levels.

To analyze the distribution of H3K23me1 in more detail, the signal was plotted along genes and TEs aligned by their transcriptional start site (TSS) and their transcriptional termination site (TTS) or their start and end coordinates, respectively. The analysis extended 2 kb upstream and downstream of the TSS and TTS, respectively, and 4 kb into the gene or TE body. Consistent with immunolocalization to pericentromeric regions, the average H3K23me1 signal was much higher along TEs than along genes (Figure 3a). H3K23me1 was present at low levels upstream and downstream of TEs. Within the first 1 kb from the border, the signal steeply increased to a maximum that remained stable in the center of the TE body. Interestingly, H3K23me1 was not restricted to heterochromatic TEs. The distribution profile along genes resembled

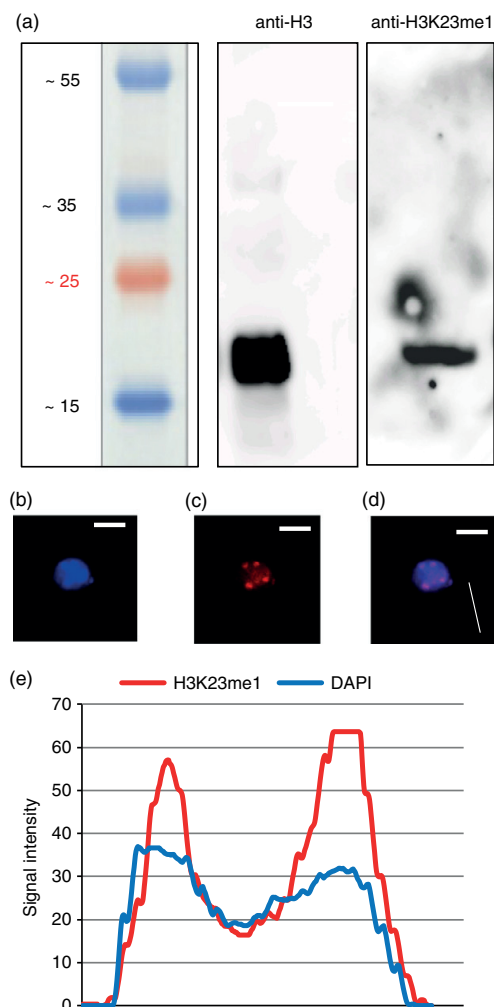


Figure 2. Identification of H3K23me1 in Arabidopsis.

(a) Acid-extracted histones from Arabidopsis inflorescences were resolved on a 15% SDS-PAGE gel, followed by protein immunoblotting using anti-H3 and anti-H3K23me1 antibodies.

(b)–(d) Root tip interphase nuclei of Arabidopsis plants were stained with 4',6-diamidino-2-phenylindole (DAPI) (b) and analyzed for immunolocalization with anti-H3K23me1 antibody (c). (d) Overlay.

(e) A section through the nucleus was analyzed using IMAGEJ software. Quantitative line profiles of DAPI (blue) and H3K23me1 (red) fluorescence intensities along the white line shown in (c). Scale bar = 5 μ m.

that along TEs but with a much lower coverage (Figure 3a).

To test if the enrichment of H3K23me1 towards the center of TEs and genes is a consequence of length, we analyzed the relation between H3K23me1 score and size. A

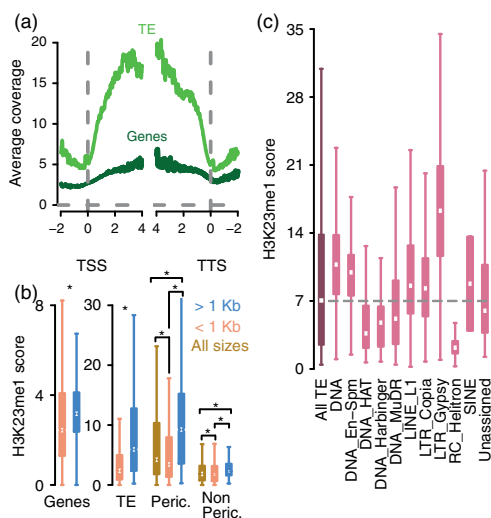


Figure 3. Genome-wide profiling of H3K23me1.

(a) Average H3K23me1 enrichment relative to the transcriptional start site (TSS) and transcriptional termination site (TTS) of genes and transposable elements (TEs). TEs were more enriched than genes. In both cases the maximum enrichment was reached towards the inside of the feature body.

(b) H3K23me1 enrichment as a function of size and pericentromeric location. Genes and TEs longer than 1 kb (blue) had a higher score than shorter ones (pink). Pericentromeric TEs (Peric.) were more enriched than non-pericentromeric (Non Peric.) ones, irrespective of their size (**P*-values from Wilcoxon test $< 2.2 \times 10^{-16}$).

(c) H3K23me1 enrichment across TE families. Members from the different TE families have a characteristic H3K23me1 signature.

H3K23me1 score was calculated as the average coverage from the 5'- to the 3'-end of the feature. Genes and TEs shorter and longer than 1 kb were analyzed separately. On average, longer genes and TEs were significantly more enriched than shorter ones (Figure 3b). Long TEs are mostly present in pericentromeric heterochromatin, whereas short TEs are often spread throughout the genome (Figure S3). To analyze if the positive correlation between size and H3K23me1 enrichment at TEs was a consequence of the chromosomal location, TEs that are located in the pericentromeric heterochromatin were analyzed separately from TEs located elsewhere in the genome. This analysis revealed that, on average, longer TEs carried more H3K23me1 than shorter ones, irrespective of their chromosomal location (Figure 3b). This shows that length and pericentromeric location can be independently associated with differential H3K23me1 enrichment. Together, the highest H3K23me1 enrichment was found preferentially on long or pericentromeric TEs.

Because long TEs had the highest H3K23me1 enrichment, we then wondered whether intrinsic characteristics derived from their sequence composition (GC content,

density of CG di-nucleotides and length) combined with components of the chromatin scaffold (H3 density, H1 presence and DNA inaccessibility) would explain H3K23me1 enrichment [H3 data from the present manuscript; H1 data from Rutowicz *et al.* (2015); DNA inaccessibility data from Shu *et al.* (2012)]. Long TEs were subdivided into 10 classes of increasing H3K23me1 enrichment and correlated with scores of predictor features. The boxplots in Figure S4(a–f) show that length, H3 score, H1 presence, CG content, DNA inaccessibility and density of CG di-nucleotides were positively correlated with H3K23me1 enrichment at long TEs. Fitting a generalized linear model using all of the predictors as explanatory variables showed that only H1 did not make a significant contribution to explaining H3K23me1 scores at long TEs (P -value = 0.02). Analysis of deviance of a reduced model not including H1 showed that length explains about 80% of the variance, while H3, CG content, DNA accessibility and density of CG di-nucleotides explain only 30% or less (Table S1). Together, intrinsic features of long TEs contributed to explain H3K23me1 scores, length being the most prominent.

Family annotation of TEs longer than 1 kb showed that H3K23me1 enrichment varied widely within and across the different families (Figure 3c). DNA, DNA-EN-Spm, LINE-L1, LTR-COPIA, SINE and LTR-Gypsy elements were more enriched than the global TE average score. Among those, LTR-Gypsy elements had the highest enrichment and the largest variation among all analyzed TE families. DNA-HAT, DNA-Harbinger and RC/Helitron elements had an average enrichment below the global TE average. RC/Helitrons had the lowest H3K23me1 score of all TE families. To further characterize the epigenetic state of the family-specific patterns of enrichment observed for H3K23me1, H3K9me2 Z-normalized scores were calculated in a similar fashion. Globally, H3K9me2 enrichment among TE families agreed with the distribution observed for H3K23me1 (Figure S5). DNA, DNA-EN-Spm, DNA-Harbinger, LINE-L1, LTR-COPIA, SINE and LTR-Gypsy TEs had a higher score than the global average. DNA-HAT, DNA-MuDR and RC/Helitron TEs scored below the average. Together, members of different TE families had a characteristic H3K23me1 enrichment that resembles H3K9me2, possibly as a consequence of the chromatin signature that regulates their potential to transpose across the genome.

H3K23me1 co-localizes with heterochromatic H3K9me2 at TEs and with euchromatic H3K4me1 at genes

Silent TEs are generally targeted by heterochromatic histone marks, whereas transcriptionally active genes are highly enriched in euchromatic marks such as H3K4me1 (Zhang *et al.*, 2009). Making use of publically available data, we investigated whether H3K23me1 and the typical euchromatic mark H3K4me1 are mutually exclusive.

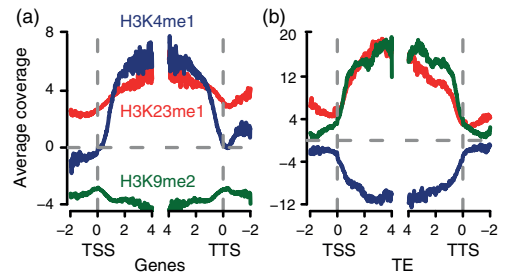


Figure 4. Co-localization of H3K23me1 with other epigenetic marks.

(a) At genes, H3K23me1 (red) co-localizes with euchromatic H3K4me1 (blue) but not with heterochromatic H3K9me2 (green). TSS, transcriptional start site; TTS, transcriptional termination site.
(b) At transposable elements (TEs), H3K23me1 co-localizes with heterochromatic H3K9me2 but not with euchromatic H3K4me1.

Unexpectedly, however, the profiles of H3K23me1 and H3K4me1 co-localized along genes (Figure 4a). This, however, was not observed for H3K4me3 or H3K4me2 (Figure S6a). Both these modifications have been found within the first 2 kb downstream of the TSS and are associated with transcription initiation (Zhang *et al.*, 2009). TEs, in contrast, did not carry any H3K4me1/2/3 but were strongly enriched in H3K23me1 (Figures 4b and S6b). The profiles of H3K36me2 and H3K36me3, two additional histone marks associated with transcription, were compared with H3K23me1 (Figure S6c, d). Profiles of both H3K36me2 and H3K36me3 enrichment differed considerably from the H3K23me1 profile. H3K36me3 has a bimodal distribution along genes. A first peak of enrichment at +629 bp is separated by a slight decrease in signal within the next 1 kb before a pronounced increase towards the central gene body (Figure S6c). H3K36me2 reached its maximum of enrichment only on the last 2 kb of the gene body (Figure S6c). H3K36me2/3 are both depleted from TEs (Figure S6d).

Consistent with earlier reports (Zhou *et al.*, 2010), the heterochromatic mark H3K9me2 was found only at TEs (Figure 4b). H3K9me2 and H3K23me1 were spatially concomitant on TEs (Figure 4b). Together, these results show that H3K23me1 resembles heterochromatic H3K9me2 along TEs but can also be found along genes, similar to the euchromatic mark H3K4me1. This suggests that H3K23me1 has a major function in heterochromatin at TEs and maybe an additional function on euchromatic genes.

KRYPTONITE affects H3K23me1 enrichment

Because of the co-localization of H3K23me1 and H3K9me2 on TEs and the enrichment of H3K23me1 in heterochromatic chromocenters (Figure 2c), we tested a potential effect of the H3K9 methyltransferase KYP on the presence of H3K23me1. To this end, we performed N-ChIP followed

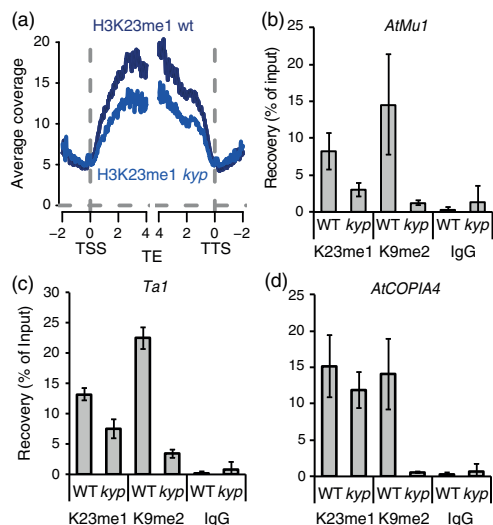


Figure 5. KRYPTONITE (KYP) is required for full H3K23me1.

(a) H3K23me1 levels along transposable elements (TEs) in the wild type (dark blue) and *kyp-6* (light blue). Enrichment of H3K23me1 was globally reduced along TEs in *kyp-6*. Data are from native chromatin immunoprecipitation sequencing. TSS, transcriptional start site; TTS, transcriptional termination site.

(b)–(d) Relative enrichment of H3K9me2 and H3K23me1 on the TE loci *AtMu1* (b), *Ta1* (*At4g03760*) (c) and *AtCOPIA4* (d) in 10-day-old wild-type (WT) and *kyp-6* seedlings. Data are from cross-linked chromatin immunoprecipitation-qPCR. Recovery with IgG is shown as a negative control. The recovery relative to input is shown as mean \pm SEM of three biological replicates.

by sequencing in the *kyp-6* mutant background. Indeed, levels of H3K23me1 along TEs were significantly reduced in the *kyp-6* mutant (Figure 5a). However, H3K23me1 was not completely depleted in the mutant background, arguing against a model where KYP is strictly required for full monomethylation of H3K23. To confirm this observation, the levels of H3K23me1 and H3K9me2 were measured by ChIP-qPCR on three known KYP targets: the transposons *Ta1* (LTR/Copia), *ATCOPIA4* (LTR/Copia) and *AtMu1* (DNA/MuDR) (Lippman *et al.*, 2003). As expected, H3K9me2 at all three tested transposons was almost completely lost in *kyp-6* (Figure 5b–d). Accumulation of H3K23me1 in the mutant was reduced by up to half of wild-type levels at the *AtMu1* and *Ta1* loci (Figure 5b, c). In contrast, H3K23me1 levels were not significantly reduced at *AtCOPIA4* (Figure 5d). This suggests that the KYP–H3K23me1 interaction is locus-dependent and confirms the requirement of KYP for full H3K23 monomethylation *in vivo*.

DNA methylation and H3K23me1 are interrelated

DNA methylation at cytosine bases is an important epigenetic modification involved, among other functions,

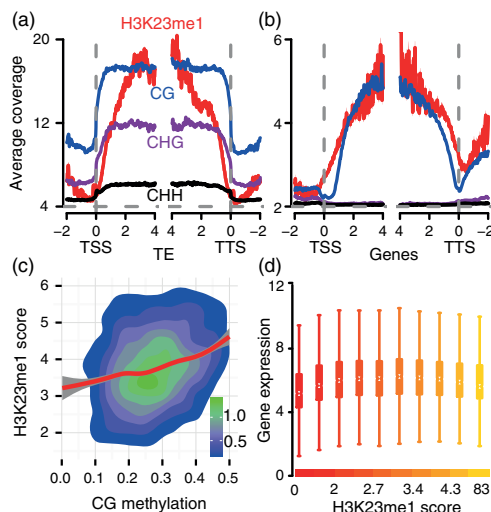


Figure 6. Co-localization of H3K23me1 and DNA methylation.

(a), (b) Average levels of H3K23me1 (red) and DNA methylation in CG (blue), CHG (purple) and CHH (black) contexts at transposable elements (TEs) (a) and genes (b). H3K23me1 co-localized with DNA methylation in the three contexts along TEs but only with CG methylation along genes. TSS, transcriptional start site; TTS, transcriptional termination site.

(c) Positive correlation of H3K23me1 enrichment and CG methylation levels in genes that carry CG methylation. Pearson correlation coefficient = 0.235, P -value $< 2.2 \times 10^{-16}$. The 95% confidence interval is depicted in light gray.

(d) Relation between H3K23me1 scores and gene expression. Genes with intermediate H3K23me1 scores were more expressed than the 10% least enriched and 10% most enriched gene classes.

in establishing heterochromatic gene silencing. The functional and mechanistic relationship between DNA methylation and H3K9me2 is well characterized in the silencing of repetitive sequences in pericentromeric regions of the Arabidopsis genome. CG, in particular, has been suggested to form a positive loop with the deposition of H3K9me2 to reinforce silencing (Caro *et al.*, 2012). To test for co-localization between H3K23me1 and DNA methylation, the average methylation profiles in CG, CHG and CHH contexts were compared with H3K23me1 enrichment. Even though H3K23me1 coexisted with DNA methylation in the three contexts along TEs, their patterns of enrichment were not similar. H3K23me1 reached a maximum only after the first 1 kb of the TE whereas DNA methylation in all three contexts is present immediately downstream of the 5'-border up to a level that remains stable until the 3'-end (Figure 6a).

It has been previously shown that in Arabidopsis genes are depleted of CHG and CHH methylation, but approximately one-third of the genes carry CG methylation in the gene body (Zilberman *et al.*, 2007). Consistent with this, H3K23me1 enrichment was highly correlated with the level

of CG methylation towards the inside of gene bodies (Figure 6b). Next, we tested whether H3K23me1 is preferentially enriched in genes with high gene-body CG methylation as defined by Zilberman *et al.* (2007). To avoid confounding effects due to bona fide heterochromatic sequences, genes covered by H3K9me2 were excluded from the analysis, leaving a total of 3532 euchromatic body-methylated genes (Table S2). CG methylation and H3K23me1 scores were calculated as the average value between TSS and TTS of annotated genes. Figure 6(c) shows that the enrichment of H3K23me1 was significantly and positively correlated with CG DNA methylation (red trend-line; Pearson correlation coefficient = 0.235, P -value < 2.2×10^{-16}).

To test whether DNA methylation can affect H3K23me1 enrichment, 10-day-old seedlings germinated on a medium supplemented with 80 μ M of the stable DNA methyltransferase inhibitor 1-(β -D-ribofuranosyl)-1,2-dihydropyrimidine-2-one (Zebularine), were used to assess possible changes in the enrichment of H3K23me1. As previously reported (Baubec *et al.*, 2009), Zebularine released TE silencing (Figure S7a, c, e), demonstrating the effectiveness of the treatment. To test the effect of the DNA hypomethylation on the level of H3K23me1, ChIP was performed on the treated and untreated samples. H3K23me1 was consistently reduced to about half at all three tested TEs (Figure S7b, d, f) suggesting that DNA methylation provides a chromatin environment that favors deposition of H3K23me1.

Enrichment of H3K23me1 does not correlate with gene expression

Because H3K23me1 was present not only on TEs but also along some genes, where it coexisted with CG gene body methylation, we wondered whether H3K23me1 was involved in transcriptional repression. To test this hypothesis, we used published transcriptome data (Shu *et al.*, 2012), which matched age, tissue and harvesting time of the H3K23me1 ChIP-seq data. The H3K23me1 enrichment was similar for all genes, for active genes, for inactive genes and for H3K27me3-silenced Polycomb group target genes (Figure S8). Features annotated as TE genes, in contrast, were clearly more enriched (Figure S8). To test if a relation of H3K23me1 and transcription existed for particular H3K23me1 enrichments, expression levels were analyzed for 10 gene classes of different H3K23me1 scores. Interestingly, the highest expression was observed for intermediate H3K23me1 scores, while both the least and the most enriched genes had lower expression levels (Figure 6d), arguing against a strong direct effect of H3K23me1 on gene expression.

H3K23me1 is a conserved modification in plants

Histones are among the most strongly conserved eukaryotic proteins. H3K23me1 was found previously in mammalian cells (Liu *et al.*, 2010b), in *C. elegans* (Vandamme *et al.*, 2015) and in the ciliate *Tetrahymena* (Zhang *et al.*, 2012). To investigate the degree of conservation of lysine 23 among other eukaryotes; a multiple sequence

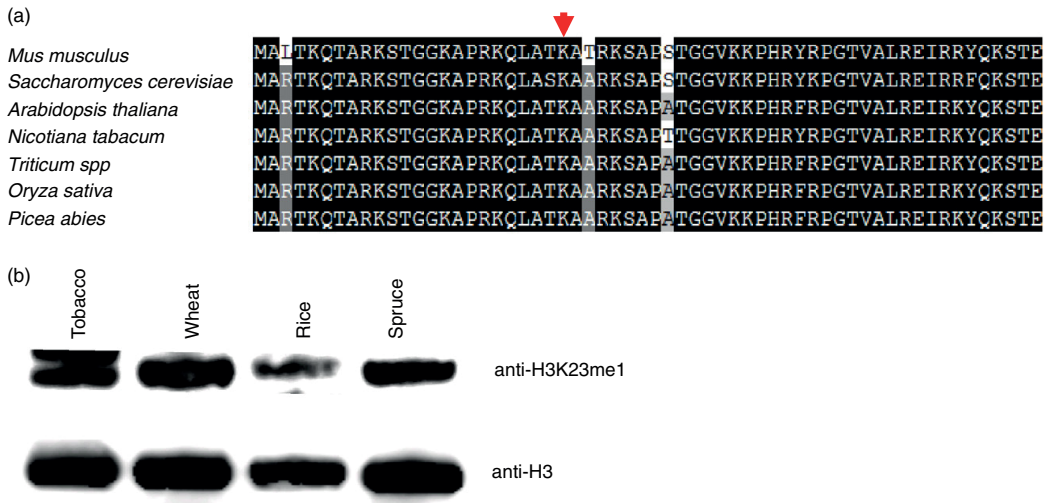


Figure 7. H3K23me1 is conserved in seed plants.

(a) Amino acid alignment of the amino-terminal tail of H3 from mouse, yeast and several plants shows conservation of lysine 23.

(b) Acid-extracted histones prepared from tobacco (*Nicotiana tabacum*), wheat (*Triticum spp*), rice (*Oryza sativa*) and spruce (*Picea abies*) were resolved on a 15% SDS-PAGE gel followed by immunoblotting using the anti-H3K23me1 antibody. Anti-H3 was used as a loading control.

alignment of the first 60 amino acids of H3 from Arabidopsis and several other representative eukaryotes was performed (Figure 7a), revealing a strong conservation of lysine 23.

To test whether H3K23me1 exists in plant species other than Arabidopsis and cauliflower, bulk histone extracts were obtained from different plant species including mono- and dicotyledonous angiosperms and gymnosperms. The presence of monomethylation at H3K23 was assessed by immunoblotting using the anti-H3K23me1 antibody. A clear band was obtained in all analyzed species (Figure 7b). These results show that H3K23me1 is a highly conserved modification in seed plants and suggest important biological functional roles for this modification.

DISCUSSION

In the present study we have established H3K23me1 at a histone H3 PTM not previously described in seed plants. H3K23me1 was detected by MS in cauliflower and by immunoblotting in cauliflower, Arabidopsis, tobacco, wheat, rice and spruce. Earlier reports described H3K23me1 in human cell lines, in *C. elegans* and in *Tetrahymena* (Liu *et al.*, 2010b; Leroy *et al.*, 2012; Zhang *et al.*, 2012; Papazyan *et al.*, 2014; Vandamme *et al.*, 2015). It remains to be tested whether H3K23me1 exists in insects, yeast and lower plants. In Arabidopsis, H3K23me1 was highly enriched in heterochromatic chromocenters and long, pericentromeric TEs. This is consistent with the observation that H3K23me1 co-localizes with the heterochromatin marker HP1 β in mammals (Liu *et al.*, 2010b) and is bound by mammalian HP1 α and HP1 β *in vivo* and *in vitro* (Liu *et al.*, 2010b; Leroy *et al.*, 2012). At the genome-wide level, it was also possible to detect H3K23me1 at a lower coverage along Arabidopsis genes. However, levels of enrichment did not correlate with gene expression. Thus, H3K23me1 does not appear to directly affect transcription rates. In both genes and TEs, the highest level of H3K23me1 was reached after the first 1000 bp of the feature body, where it remained stable before steeply decreasing towards the 3' end. At the regions of higher H3K23me1 enrichment, it co-localized with the euchromatic mark H3K4me1 in genes and with the heterochromatic mark H3K9me2 in TEs. In *C. elegans*, H3K23me1 was found together with H3K27me3 on silent PcG target genes (Vandamme *et al.*, 2015); however, we did not see this in Arabidopsis (Figure S8). Different TE families in Arabidopsis have a distinctive enrichment of H3K23me1. In particular, LTR-Gypsy TEs have the highest H3K23me1 score among all families. Similar patterns were observed for H3K9me2. The characteristically high levels of H3K23me1 in those elements suggest that H3K23me1 is part of the chromatin state that reinforces constitutive silencing by H3K9me2. Members of the RC/Helitron family, on the other hand, have the lowest H3K23me1 and H3K9me2 scores.

Helitrons, as actively mutagenic members, may be able to avoid the epigenetic silencing mechanisms in order to retain the potential to transpose across the genome.

In Arabidopsis heterochromatin is marked by H3K9me2 (Jackson *et al.*, 2004), which is deposited by the HKMT KYP (Jackson *et al.*, 2004). We tested the relation between H3K9me2 and H3K23me1 using a *kyp* mutant. Globally, levels of H3K23me1 were reduced in TEs. Measurement of the levels of H3K23me1 and H3K9me2 in silenced loci showed that, in contrast to H3K9me2 which was almost entirely lost in *kyp-6*, H3K23me1 was reduced to only about 50%. Thus, KYP is needed for full H3K23me1 levels but is not essential for this modification. It is possible that presence of H3K9me2 favors the recruitment of an unknown HKMT that adds H3K23me1 to local chromatin.

In addition to H3K9me2, DNA methylation in all three sequence contexts is another marker of heterochromatin in Arabidopsis (He *et al.*, 2011). In TEs, H3K23me1 was concomitant with the three different contexts of DNA methylation, whereas along genes it only co-localized with CG methylation. Analysis of the genes that carry CG methylation in the gene body showed a positive correlation between the two marks. These results suggest that H3K23me1 deposition may be related to DNA methylation. To further test whether DNA methylation can recruit H3K23me1, we used an inhibitor of the catalytic activity of DNA methyl transferases. Because the commonly used cytidine analog 5-azacytidine is relatively unstable and often toxic, the more stable and less toxic inhibitor Zebularine is preferred as an efficient means to block DNA methylation (Zhou *et al.*, 2002; Cheng *et al.*, 2003). This drug has been successfully used in mammals, fungi, ciliates and plants (Cheng *et al.*, 2003; Baubec *et al.*, 2009; Malik *et al.*, 2012). We used Zebularine to test a potential relationship between H3K23me1 and DNA methylation. Indeed, H3K23me1 was reduced upon Zebularine treatment at the analyzed loci. This suggests that DNA methylation favors monomethylation of H3K23me1. It remains to be tested whether H3K23me1 in turn can recruit DNA methyltransferases as previously shown for H3K9me2 (Du *et al.*, 2012). Together, H3K23me1 is enriched in heterochromatin in Arabidopsis and depends, at least partially, on undisturbed H3K9me2 and DNA methylation.

EXPERIMENTAL PROCEDURES

Plant material and growth conditions

The *Arabidopsis thaliana* (L.) Heynh wild-type accession Columbia (Col), tobacco (*Nicotiana tabacum*), wheat (*Triticum* spp.), rice (*Oryza sativa*) and Norway spruce (*Picea abies*) were used. The *kyp-6* allele (SALK_041474) (Chan *et al.*, 2006) was obtained from the Nottingham Seed Stock Centre. Seeds were sown on 0.5 \times basal salts Murashige and Skoog (MS) medium (Duchefa, <https://www.duchefa-biochemie.com/>), stratified at 4°C for 1 day and then transferred to growth chambers for germination at 20°C

under long-day (LD, 16-h light) photoperiods for 10 days before being transferred to soil. Wheat and rice were grown in a greenhouse at 22°C and LD. Norway spruce needles were from a wild-grown tree close to Uppsala (Sweden). For treatments with the DNA methylation inhibitor Zebularine (AH Diagnostics, <http://www.ahdiagnostics.com/>), sterilized Arabidopsis seeds were germinated on MS medium containing 80 μM Zebularine. After 10 days, seedlings were collected for RNA extraction and ChIP assays.

RNA isolation and RT-qPCR

RNA extraction and reverse transcription were performed as described previously (Alexandre *et al.*, 2009) with minor modifications. DNA-free RNA was reverse-transcribed using a RevertAid First Strand cDNA Synthesis Kit (Fermentas, <https://www.thermofisher.com/>). Quantitative real-time PCR (qPCR) was performed with gene-specific primers (Table S3) using SYBR green (Fermentas) on an IQ5 multicolor Real time PCR cyclyer (Bio-Rad, <http://www.bio-rad.com/>). The qPCR reactions were performed in triplicate, gene expression levels were normalized to a PP2A control gene and results were analyzed as described (Simon, 2003).

Chromatin immunoprecipitation and ChIP-seq

Cross-linked chromatin immunoprecipitation (X-ChIP) assays were performed as described (Exner *et al.*, 2009). Antibodies used for ChIP were anti-H3 (Upstate/Millipore, <http://www.merckmillipore.com/>; catalog no. 07-690), anti-H3K9me2 (Diagenode, <https://www.diagenode.com/>; catalog no. pAb-060-050), anti-H3K23me1 (ActiveMotif, <https://www.activemotif.com/>; catalog no. 39387) and non-immune IgG (Sigma-Aldrich, <http://www.sigmaaldrich.com/>; catalog no. I5006). Quantitative real-time PCR using SYBR green (Fermentas) and gene-specific primers (Table S4) was performed on an IQ5 multicolor Real time PCR cyclyer (Bio-Rad). Results were normalized to input and presented as the enrichment level of the given modification over the level of histone H3. Chromatin immunoprecipitation was performed in two biological replicates.

Native chromatin immunoprecipitation was performed following the procedure of Mahrez *et al.* (2016) with slight modifications, using 400 mg of fifth and sixth rosette leaves of 5-week-old plants grown under short-day conditions. Tissue was homogenized in Honda buffer [2.5% w/v Ficoll 400, 5% dextran T40, 0.4 M sucrose, 25 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (TRIS)-HCl pH 7.4, 10 mM MgCl_2 , 10 mM β -mercaptoethanol and 0.5% Triton X-100] using a GentleMACS dissociator (Miltenyi Biotec, <http://www.miltenyibiotec.com/en/>). The homogenate was incubated for 15 min with gentle rotation, followed by filtering twice through Miracloth and once through 30 μm CellTrics (Sysmex, Kobe, Japan). The nuclear suspension was pelleted at 1500 g for 6 min at 4°C. Pellets were washed once with 1 ml of MNase buffer [50 mM TRIS pH 8, 10 mM NaCl, 5 mM CaCl_2 and EDTA-free protease inhibitor cocktail (Roche, <http://www.roche.com/>)] and resuspended in 100 μl of MNase buffer with RNase A (EN0531, Thermo Scientific, <http://www.thermofisher.com/>). Isolated nuclei were treated with 2 μl of Micrococcal Nuclease (13.3 units μl^{-1} , Fermentas) for 7 min at 37°C. The reaction was stopped with EDTA (final concentration 10 mM) and incubated for 10 min on ice. The supernatant after centrifugation (2000 g , 4°C, 5 min) was collected and the pellet was resuspended in 100 μl of S2 buffer (1 mM TRIS pH 8, 0.2 mM EDTA and EDTA-free protease inhibitor cocktail). After incubation on ice for 30 min, the supernatant was combined with the first collected and 200 μl of N-ChIP 2 \times salt buffer (25 mM TRIS pH 8, 100 mM NaCl, 5 mM EDTA, 0.2% Triton X-100 and EDTA-free protease inhibitor cocktail) were added. The chromatin was obtained from the

supernatant obtained after centrifugation at 17 000 g and 4°C. Ten microliters was kept as Input and a minimum of 100 μl of chromatin was incubated with antibodies at 4°C with rotation overnight. Protein-A Dynabeads (Invitrogen, <http://www.invitrogen.com/>) were added followed by continued incubation for 90 min. Beads were washed once briefly and once after a 10-min rotation at 4°C with each of three washing buffers with increasing NaCl concentration (washing buffers: 25 mM TRIS pH 8, 10 mM EDTA, 50/100/150 mM NaCl, 0.1% Triton X-100 and EDTA-free protease inhibitor cocktail). A final short wash was done with Tris-EDTA buffer (pH 8) and beads were immediately collected. DNA elution, de-crosslinking and purification steps were done using the IPure kit (Diagenode) following the manufacturer's protocol. The following antibodies were used: anti-Histone H3 (Sigma, catalog no. H9289) and anti-H3K23me1 (ActiveMotif, catalog no. 39387). Quantitative real-time PCR using SYBR green (Fermentas) and gene-specific primers (Table S4) was performed on an IQ5 multicolor Real time PCR cyclyer (Bio-Rad) as a quality control before library preparation. Biological duplicates were used for sequencing. Libraries were generated using the Ovation Ultralow Library System (NuGEN, <http://www.nugen.com/>) following the manufacturer's protocol using 1 ng of starting material previously sonicated for 25 cycles (30 sec ON, 30 sec OFF at high power) using a Bioruptor (Diagenode). Sequencing was performed on an Illumina HiSeq2000 in 50-bp single-end mode.

Immunostaining

Immunostaining was performed as described (Jasencakova *et al.*, 2000) with some modifications. Arabidopsis root tips were pre-fixed with paraformaldehyde, washed in MTBS buffer [50 mM piperazine-*N,N*-bis(2-ethanesulfonic acid) (PIPES), 5 mM MgSO_4 , and 5 mM EGTA, pH 7.9], and digested for 10 min at 37°C with a PCP enzyme mixture [2.5% pectinase, 2.5% cellulase Onozuka R-10, and 2.5% Pectolyase Y-23 (w/v) dissolved in MTBS]. After washing with MTBS buffer, root tips were squashed and frozen in liquid nitrogen. Nuclei were covered with blocking solution (MTBS containing 3% BSA) for 1 h at 4°C, incubated for 1 h at 4°C with anti-H3K23me1 (ActiveMotif, catalog no. 39387) in MTBS (containing 1% BSA and 0.1% Tween 20) and kept in a humid chamber overnight. After washing (three times with MTBS for 5 min each), slides were covered for 2 h at 4°C with Rhodamine-conjugated anti-rabbit IgG antibody (Thermo Scientific) diluted in MTBS containing 1% BSA and 0.1% Tween 20. Slides were gently washed with MTBS, and DNA was stained with 1 $\mu\text{g ml}^{-1}$ of DAPI (Vectashield; Vector Laboratories, <https://vectorlabs.com/>). Fluorescence signal detection and documentation was performed with a Leica DMI 4000 microscope.

Histone extraction from plants, fractionation by reverse-phase HPLC and identification of PTMs by liquid chromatography mass spectrometry

Histone extraction from plants, fractionation by reverse-phase HPLC and PTM identification by liquid chromatography mass spectrometry (LC-MS/MS) were described previously (Mahrez *et al.*, 2016).

Sequence alignment of H3 amino-terminal tails

Amino acid sequences of H3 from representative eukaryotes were selected using PSI-BLAST searches. Sequences were aligned using the CLUSTALW multiple sequence alignment program implemented in MEGA5 (Tamura *et al.*, 2011). Visualization of the sequence alignment was performed with GenDoc; the amino acid sequences are available in the FASTA format in Table S5.

Bioinformatics and ChIP-Seq data analysis

Reads were mapped to the Arabidopsis reference TAIR10 genome using bowtie2 (version 2.1; Langmead and Salzberg, 2012). Details on read numbers can be found in Table S6. SAM file output from bowtie2 was converted to BAM format using SAMtools (version 1.4; Li et al., 2009) and imported into R (version 2.15.2; <http://www.R-project.org/>) using functions from the Rsamtools package. All subsequent analysis was performed in R. Identical reads present more than 25 times were considered as PCR artifacts and filtered out using the filterDuplReads function from package HtSeqTools. The two sequenced replicates were pooled. Each sample was normalized by the total number of sequenced reads. H3K23me1 signal was normalized to histone signal to control for variable nucleosome density using functions from the package nucleR (Flores and Orozco, 2011). The pericentromeric heterochromatin was considered to span the regions between the following coordinates: Chr1, 11 500 020–17 696 331; Chr2, 1 100 003–7 192 918; Chr3, 10 298 763–17 289 015; Chr4, 1 500 001–2 300 002; Chr4, 2 800 003–6 300 004; Chr5, 8 999 997–5 982 772 (Copenhaver et al., 1999). A gene score was defined as the average coverage between the TSS and the TTS. The density of CG dinucleotides refers to the total number of CGs along a given TE normalized by its length. The following data were taken from the literature: H3K4me1 (Zhang et al., 2009); H3K9me2 (Rehrauer et al., 2010); H3K4me2/3 and H3K36me2 (Luo et al., 2013); H3K36me3 (Mahrez et al., 2016); H1 (Rutowicz et al., 2015); DNA methylation data (Zemach et al., 2013). For the metagene plots, the coverage values of H3K4me1/2/3, H3K36me2/3 and H3K9me2 were scaled for plotting together with H3K23me1. Gene expression data are from Shu et al. (2012). The expression cutoff to consider a gene as active was a value of 6 in the microarray data from Shu et al. (2012), resulting in a set of 12 219 active protein-coding genes. The other 14 563 genes were considered as inactive. Genes carrying CG methylation in the gene body were taken from Zilberman et al. (2007) and filtered to remove genes carrying H3K9me2 or TEs.

ACCESSION NUMBERS

Sequencing reads are deposited in the Gene Expression Omnibus (GSE86498).

ACKNOWLEDGMENTS

We are very grateful to Drs Sun Chuanxin, Joel Sohlberg and Mohammad Sameri for kindly providing seeds for rice and leaf material for spruce and wheat, respectively. This work was supported by grants from the Swiss National Science Foundation SNF and the Swedish Science Foundation VR, the Knut-and-Alice-Wallenberg Foundation and the Carl-Tryggers-Foundation. The authors have no conflicts of interest to declare.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Identification of H3K23ac.

Figure S2. Genome-wide distribution of H3K23me1 on the Arabidopsis genome.

Figure S3. Proportion of short and long transposable elements in pericentromeric and non-pericentromeric regions.

Figure S4. Intrinsic features of long transposable elements partially explained H3K23me1 enrichment.

Figure S5. H3K9me2 enrichment in transposable element families.

Figure S6. Co-localization of H3K23me1 with other epigenetic marks of active chromatin.

Figure S7. DNA methylation and H3K23me1 are interrelated epigenetic marks.

Figure S8. Enrichment of H3K23me1 does not correlate with gene expression.

Table S1. Generalized linear model for H3K23me1 scores at long transposable elements.

Table S2. List of active genes that carry CG gene body methylation as defined by Zilberman et al. (2007) after removing genes with H3K9me2.

Table S3. List of gene-specific primers used for RT-qPCR assays.

Table S4. List of gene-specific primers used for chromatin immunoprecipitation-qPCR assay.

Table S5. The sequences of the 60 amino-terminal amino acids of histone H3.1 from different organisms.

Table S6. Summary of chromatin immunoprecipitation sequencing experiments.

REFERENCES

- Alexandre, C., Moller-Steinbach, Y., Schonrock, N., Grisseum, W. and Hennig, L. (2009) Arabidopsis MSI1 is required for negative regulation of the response to drought stress. *Molecular Plant*, **2**, 675–687.
- Bannister, A.J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*, **21**, 381–395.
- Baubec, T., Pecinka, A., Rozhon, W. and Mittelsten Scheid, O. (2009) Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *Plant J.*, **57**, 542–554.
- Bernatavichute, Y.V., Zhang, X.Y., Cokus, S., Pellegrini, M. and Jacobsen, S.E. (2008) Genome-Wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS ONE* **3**(9), e3156.
- Black, J.C., Van Rechem, C. and Whetstone, J.R. (2012) Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell*, **48**, 491–507.
- Caro, E., Stroud, H., Greenberg, M.V., Bernatavichute, Y.V., Feng, S., Groth, M., Vashisht, A.A., Wohlschlegel, J. and Jacobsen, S.E. (2012) The SET-domain protein SUVH5 mediates H3K9me2 deposition and silencing at stimulus response genes in a DNA methylation-independent manner. *PLoS Genet.*, **8**, e1002995.
- Chan, S.W., Henderson, I.R., Zhang, X., Shah, G., Chien, J.S. and Jacobsen, S.E. (2006) RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in Arabidopsis. *PLoS Genet.*, **2**, e83.
- Charron, J.B.F., He, H., Elling, A.A. and Deng, X.W. (2009) Dynamic landscapes of four histone modifications during deetiolation in Arabidopsis. *Plant Cell*, **21**, 3732–3748.
- Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S.C., Falck, J.R., Peng, J., Gu, W. and Zhao, Y. (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell Proteomics*, **6**, 812–819.
- Cheng, J.C., Matsen, C.B., Gonzales, F.A., Ye, W., Greer, S., Marquez, V.E., Jones, P.A. and Selker, E.U. (2003) Inhibition of DNA methylation and reactivation of silenced genes by zebularine. *J Natl Cancer I*, **95**, 399–409.
- Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., Stallcup, M.R. and Cheng, X.D. (2008) The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat. Struct. Mol. Biol.*, **15**, 245–250.
- Copenhaver, G.P., Nickel, K., Kuromori, T. et al. (1999) Genetic definition and sequence analysis of Arabidopsis centromeres. *Science*, **286**, 2468.
- Du, J., Zhong, X., Bernatavichute, Y.V. et al. (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167–180.
- Exner, V., Aichinger, E., Shu, H., Wildhaber, T., Alfaro, P., Cafisch, A., Grisseum, W., Kohler, C. and Hennig, L. (2009) The chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 is essential for H3K27me3 binding and function during Arabidopsis development. *PLoS ONE*, **4**, e5335.
- Flores, O. and Orozco, M. (2011) nucleR: a package for non-parametric nucleosome positioning. *Bioinformatics*, **27**, 2149–2150.

- Fransz, P., ten Hoopen, R. and Tessoro, F. (2006) Composition and formation of heterochromatin in *Arabidopsis thaliana*. *Chromosome Res.* **14**, 71–82.
- Fujiki, R., Hashiba, W., Sekine, H. *et al.* (2011) GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature*, **480**, 557–560.
- Greer, E.L. and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–357.
- He, X.J., Chen, T.P. and Zhu, J.K. (2011) Regulation and function of DNA methylation in plants and animals. *Cell Res.* **21**, 442–465.
- Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B. and Aasland, R. (2011) The CW domain, a new histone recognition module in chromatin proteins. *EMBO J.* **30**, 1939–1952.
- Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., Cheng, X.D., Schubert, I., Jenuwein, T. and Jacobsen, S.E. (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma*, **112**, 308–315.
- Jasencakova, Z., Meister, A., Walter, J., Turner, B.M. and Schubert, I. (2000) Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. *Plant Cell*, **12**, 2087–2100.
- Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
- Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J. and Jacobsen, S.E. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* **17**, 379–384.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220.
- Leroy, G., Chepelev, I., Dimaggio, P.A., Blanco, M.A., Zee, B.M., Zhao, K. and Garcia, B.A. (2012) Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. *Genome Biol.* **13**, R68.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Lippman, Z., May, B., Yordan, C., Singer, T. and Martienssen, R. (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1**, e67.
- Liu, C.Y., Lu, F.L., Cui, X. and Cao, X.F. (2010a) Histone methylation in higher plants. *Annu. Rev. Plant Biol.* **61**, 395–420.
- Liu, H., Galka, M., Iberg, A. *et al.* (2010b) Systematic identification of Methyllysine-Driven interactions for histone and nonhistone targets. *J. Proteome Res.* **9**, 5827–5836.
- Loidl, P. (2004) A plant dialect of the histone language. *Trends Plant Sci.* **9**, 84–90.
- Luo, C., Sidote, D.J., Zhang, Y., Kerstetter, R.A., Michael, T.P. and Lam, E. (2013) Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. *Plant J.* **73**, 77–90.
- Mahrez, W., Arellano, M.S., Moreno-Romero, J., Nakamura, M., Shu, H., Nanni, P., Kohler, C., Grissem, W. and Hennig, L. (2016) H3K36ac is an evolutionary conserved plant histone modification that marks active genes. *Plant Physiol.* **170**, 1566–1577.
- Malik, G., Dangwal, M., Kapoor, S. and Kapoor, M. (2012) Role of DNA methylation in growth and differentiation in *Physcomitrella patens* and characterization of cytosine DNA methyltransferases. *FEBS J.* **279**, 4081–4094.
- Martin, C. and Zhang, Y. (2005) The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Bio.* **6**, 838–849.
- Papayan, R., Voronina, E., Chapman, J.R. *et al.* (2014) Methylation of histone H3K23 blocks DNA damage in pericentric heterochromatin during meiosis. *Elife*, **3**, e02996.
- Rehauer, H., Aquino, C., Grissem, W. *et al.* (2010) AGRONOMICS1: a new resource for *Arabidopsis* transcriptome profiling. *Plant Physiol.* **152**, 487–499.
- Roudier, F., Ahmed, I., Berard, C. *et al.* (2011) Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* **30**, 1928–1938.
- Rutowicz, K., Puzio, M., Halibart-Puzio, J. *et al.* (2015) A specialized histone H1 variant is required for adaptive responses to complex abiotic stress and related DNA methylation in *Arabidopsis*. *Plant Physiol.* **169**, 2080–2101.
- Shu, H., Wildhaber, T., Siretskiy, A., Grissem, W. and Hennig, L. (2012) Distinct modes of DNA accessibility in plant chromatin. *Nat. Commun.* **3**, 1281.
- Simon, P. (2003) Q-Gene: processing quantitative real-time RT-PCR data. *Bioinformatics*, **19**, 1439–1440.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tan, M., Luo, H., Lee, S. *et al.* (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, **146**, 1016–1028.
- Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D. and Patel, D.J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040.
- Vandamme, J., Sidoli, S., Mariani, L., Friis, C., Christensen, J., Helin, K., Jensen, O.N. and Salcini, A.E. (2015) H3K23me2 is a new heterochromatic mark in *Caenorhabditis elegans*. *Nucleic Acids Res.* **43**, 9694–9710.
- Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L. and Zilberman, D. (2013) The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, **153**, 193–205.
- Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M. and Jacobsen, S.E. (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **10**, 1–14.
- Zhang, C., Molascon, A.J., Gao, S., Liu, Y. and Andrews, P.C. (2012) Quantitative proteomics reveals that the specific methyltransferases Txr1p and Etl2p differentially affect the mono-, di- and tri-methylation states of histone H3 lysine 27. *Mol. Cell Proteomics*, **12**(6), 1678–1688.
- Zhou, D.X. (2009) Regulatory mechanism of histone epigenetic modifications in plants. *Epigenetics-Us*, **4**, 15–18.
- Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J. and Hornby, D.P. (2002) Zebularine: A novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J. Mol. Biol.* **321**, 591–599.
- Zhou, J., Wang, X., He, K., Charron, J.-B.F., Elling, A.A. and Deng, X.W. (2010) Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. *Plant Mol. Biol.* **72**, 585–595.
- Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. and Henikoff, S. (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**, 61–69.

Arabidopsis Chromatin Assembly Factor 1 is required for occupancy and position of a subset of nucleosomes

Rafael Muñoz-Viana¹, Thomas Wildhaber², Minerva S. Trejo-Arellano¹ , Iva Mozgová^{1,3}  and Lars Hennig^{1,*} 

¹Department of Plant Biology and Linnean Center for Plant Biology, Swedish University of Agricultural Sciences, PO-Box 7080, SE-75007 Uppsala, Sweden,

²Department of Biology, ETH Zürich, Universitätsstrasse 2, CH-8092 Zürich, Switzerland, and

³Institute of Microbiology, Centre Algatech, Opatovický mlýn, 37981 Treboň, Czech Republic

Received 20 January 2017; revised 21 July 2017; accepted 1 August 2017; published online 8 August 2017.

*For correspondence (e-mail lars.hennig@slu.se).

SUMMARY

Chromatin Assembly Factor 1 (CAF-1) is a major nucleosome assembly complex which functions particularly during DNA replication and repair. Here we studied how the nucleosome landscape changes in a CAF-1 mutant in the model plant *Arabidopsis thaliana*. Globally, most nucleosomes were not affected by loss of CAF-1, indicating the presence of efficient alternative nucleosome assemblers. Nucleosomes that we found depleted in the CAF-1 mutant were enriched in non-transcribed regions, consistent with the notion that CAF-1-independent nucleosome assembly can compensate for loss of CAF-1 mainly in transcribed regions. Depleted nucleosomes were particularly enriched in proximal promoters, suggesting that CAF-1-independent nucleosome assembly mechanisms are often not efficient upstream of transcription start sites. Genes related to plant defense were particularly prone to lose nucleosomes in their promoters upon CAF-1 depletion. Reduced nucleosome occupancy at promoters of many defense-related genes is associated with a primed gene expression state that may considerably increase plant fitness by facilitating plant defense. Together, our results establish that the nucleosome landscape in *Arabidopsis* is surprisingly robust even in the absence of the dedicated nucleosome assembly machinery CAF-1 and that CAF-1-independent nucleosome assembly mechanisms are less efficient in particular genome regions.

Keywords: *Arabidopsis thaliana*, chromatin, CAF-1, plant defense, GSE87421.

INTRODUCTION

Nuclear DNA is packaged with histone proteins into nucleosomes forming the basic unit of chromatin. Nucleosome formation is catalyzed by replication-dependent and -independent mechanisms. Chromatin assembly factor 1 (CAF-1) is the major chromatin assembler for H3–H4 during replication (Stillman, 1986; Smith and Stillman, 1989). Direct interaction with proliferating cell nuclear antigen (PCNA) targets CAF-1 to sites of replication (Shibahara and Stillman, 1999). CAF-1 is conserved in all eukaryotes and consists of three subunits called p150, p60 and p48 in mammals (Kaufman *et al.*, 1995) and FASCIATA 1 (FAS1), FASCIATA 2 (FAS2) and MULTICOPY SUPPRESSOR OF IRA (MSI1) in plants (Kaya *et al.*, 2001). The two larger subunits FAS1/p150 and FAS2/p60 function exclusively in CAF-1 while the smallest subunit MSI1/p48 is a histone chaperone that functions in diverse complexes involved in chromatin formation, modification or remodeling (Hennig *et al.*, 2005). CAF-1 binds to two H3–H4 dimers and promotes formation

of (H3–H4)₂ tetramers (Winkler *et al.*, 2012). Animals and plants possess specific histone H3 variants for replication-dependent and -independent chromatin assembly (Talbert and Henikoff, 2010); and human CAF-1 binds *in vivo* to the replication-specific histone variant H3.1 but not to the replication-independent histone variant H3.3 (Tagami *et al.*, 2004). In contrast, the HIRA histone chaperone normally binds mostly to H3.3 but can bind H3.1 in cells depleted of CAF-1 (Tagami *et al.*, 2004; Ray-Gallet *et al.*, 2011). Also in plants and yeast, HIRA partially compensates for loss of CAF-1 (Kaufman *et al.*, 1998; Duc *et al.*, 2015).

CAF-1 is essential in animals and causes delays in the cell cycle and increased UV-sensitivity in yeast (Kaufman *et al.*, 1997; Hoek and Stillman, 2003; Nabatiyan and Krude, 2004; Houlard *et al.*, 2006; Song *et al.*, 2007). CAF-1 function is required for normal development of multicellular organisms (Yu *et al.*, 2015). In the model plant *Arabidopsis thaliana* (Arabidopsis), the absence of a functional CAF-1

complex is not lethal, as *fas1* and *fas2* mutants are viable (Reinholz, 1966; Leyser and Furner, 1992; for review see Ramirez-Parra and Gutierrez, 2007b). Mutants for the smallest subunit, MSI1, are embryo-lethal but this is CAF-1-independent and results from the function of MSI1 in Polycomb repressive complex 2 (Köhler *et al.*, 2003; Guitton *et al.*, 2004). Mutants for CAF-1 in Arabidopsis show several phenotypic defects, including stem fasciation, abnormal leaf and flower morphology, and disorganization of the shoot and root apical meristems, but are viable (Reinholz, 1966; Leyser and Furner, 1992; Kaya *et al.*, 2001). Also in rice, CAF-1 is required for normal development of the shoot apical meristem (SAM) (Abe *et al.*, 2008). In contrast to the increased size of the SAM in Arabidopsis CAF-1 mutants, rice CAF-1 mutants have a reduced SAM, indicating that mechanisms of SAM maintenance differ between rice and Arabidopsis. Additionally, Arabidopsis CAF-1 affects heterochromatin, transcriptional gene silencing, endoreduplication, cell differentiation, cell cycle duration, homologous recombination and trichome development (Kaya *et al.*, 2001; Endo *et al.*, 2006; Exner *et al.*, 2006, 2008; Kirik *et al.*, 2006; Ono *et al.*, 2006; Schönrock *et al.*, 2006; Ramirez-Parra and Gutierrez, 2007a; Abe *et al.*, 2008; Chen *et al.*, 2008). Loss of Arabidopsis CAF-1 causes also telomere shortening and reduction of ribosomal DNA clusters (Mozgova *et al.*, 2010; Muchova *et al.*, 2015; Havlova *et al.*, 2016; Pavlistova *et al.*, 2016).

The viability of plant CAF-1 mutants clearly demonstrates that compensating nucleosome assembly mechanisms exist, likely involving HIRA (Duc *et al.*, 2015). This is consistent with results from mammalian cells where HIRA-dependent nucleosome gap filling was observed (Ray-Gallet *et al.*, 2011). However, it is so far unknown how the genome-wide nucleosome landscape differs between wild-type and CAF-1-deficient cells. There are several important open questions. Is there a genome-wide loss of nucleosomes in CAF-1 mutants? Are there local regions with reduced nucleosome occupancy? Are changes a consequence of altered transcription? Here we describe experiments involving micrococcal nuclease (MNase) digestion followed by sequencing (MNase-seq) that establish specific nucleosome landscapes in a CAF-1 mutant. We find that global nucleosome patterns are mostly unchanged but that specific nucleosomes are depleted or shifted. Depleted nucleosomes are often in promoters, particularly in proximal promoters of defense-related genes.

RESULTS AND DISCUSSION

Genome sequence loss is not widespread in the *fas2* CAF-1 mutant

MNase-seq is commonly used to obtain high-resolution genome-wide maps of nucleosomes (Henikoff *et al.*, 2011). Here, we applied MNase-seq to compare the nucleosome

landscape in the Arabidopsis CAF-1 mutant *fas2-4* with that in wild-type. Arabidopsis CAF-1 mutants that have lost certain sequences from their genome, such as 45S rDNA sequences (Mozgova *et al.*, 2010), which could interfere with the interpretation of MNase-seq data. Because it was not known whether other genome regions also undergo selective loss in *fas2*, we first used genome resequencing to probe for loss of genome regions in *fas2* plants of generation 6. First, we compared sequence coverage between the wild type and *fas2* for 45S rDNA sequences, which are known to be lost in *fas2*, and for other tandem or dispersed repetitive sequences, which are stable in *fas2* (Mozgova *et al.*, 2010). Consistent with the earlier report, about 90% of the 45S rDNA reads were lost in *fas2* (Table S1 in the Supporting Information). In contrast, the tandem 5S rDNA repeats, the 180-bp centromeric repeats (Murata *et al.*, 1994) dispersed repeats, such as the retrotransposon *ATHILA* or the transposon *CACTA1*, as well as the single-copy controls *PP2A* and *ACTIN* had very similar coverage in the wild type and the mutant, arguing against loss of these sequences in *fas2*. Next, we divided the annotated genome into 1-kb bins and tested which regions have reduced coverage in *fas2*. Four such genome regions were identified, comprising in total 27 kb (Table 1). Two regions were long (10 kb and 15 kb) while the other regions were individual 1-kb fragments. The 10-kb region with reduced coverage in *fas2* includes transposable element (TE)-rich 45S rDNA-containing sequences bordering Nucleolus Organizer Region 2 (NOR2). Interestingly, the region bordering NOR4, which does not contain any 45S rDNA annotated units, was not detected as changed. The 15-kb region with reduced coverage in *fas2* is a pericentromeric TE-rich region on chromosome 3 that also contains annotated 5.8S and 18S rRNA genes, which was not known to be affected in *fas2*. The other regions with reduced coverage in *fas2* were located in genes. We note that the reference genome-based resequencing analysis can be sensitive to errors in the TAIR10 genome assembly. Nevertheless, the experiment established that genome sequence loss in *fas2* is not widespread but restricted to sequences located in a few selected regions, in particular 45S rDNA and a pericentromeric TE-rich, rDNA-containing region on chromosome 3. Independent experiments using quantitative PCR on a subset of genomic loci confirmed the results of the sequencing for all but one of the individual 1-kb fragment regions (Figure 1). All regions with reduced coverage in *fas2* were excluded from the analysis of nucleosome occupation.

Nucleosomes with reduced occupancy in the *fas2* CAF-1 mutant are enriched in non-transcribed genome regions

In contrast to the assumption that all nucleosomes are similarly digested by MNase, nucleosomes containing particular histone variants are much more sensitive to this

Table 1 Genome regions with reduced sequence coverage in *fas2*. Sequence coverage in Col and *fas2* was compared for bins of 1000 bp. Regions (single bins or merged adjacent significant bins) with a significantly (one-sided binomial test, multiple testing correction according to Storey and Tibshirani (2003); $q < 0.05$) reduced coverage in *fas2* are listed

Chromosome	Start	End	Counts in Col	Counts in <i>fas2</i>	Log ₂ FC	Comment
Chr2	1001	11 000	188 137	22 542	−3	NOR2 border region ^a
Chr2	7 191 001	7 192 000	5551	625	−3	<i>AT2G16586</i>
Chr3	14 191 001	14 206 000	3069	449	−3	Pericentromeric region ^b
Chr5	3 253 001	3 254 000	16 699	1833	−3	Upstream ^c of <i>AT5G10340</i>

FC, fold change.

^aIncludes: *AT2G01008*, *AT2G01010* (18S rRNA), *AT2G01020* (5.8S rRNA), *AT2G01021*, *AT2G01023*, *AT2G01022*.

^bIncludes: *AT3G41761*, *AT3G41762*, *AT3G41768* (18S rRNA), *AT3G41979* (5.8S rRNA).

^cUpstream: within 1000 bp from the 5' end of the gene.

treatment (Xi *et al.*, 2011). H3.3- or H2A.Z-containing nucleosomes, for instance, are more fragile (Jin and Felsenfeld, 2007; Jin *et al.*, 2009). It has been shown that chemical cross-linking greatly suppresses the high sensitivity of most unstable nucleosomes to MNase (Xi *et al.*, 2011). Here, we used MNase digestion of cross-linked chromatin to reduce the loss of fragile nucleosomes. DNA from duplicates of MNase-digested cross-linked chromatin of Col wild type and *fas2* was subjected to Illumina sequencing. For chromatin extraction, leaf number 6 was harvested from 38-day-old plants. At this developmental stage, leaf number 6 is fully expanded and all cell division and endoreduplication has ceased (Baerenfaller *et al.*, 2012). Observed changes in nucleosomes are therefore not transient defects present only during or shortly after replication but are more permanent chromatin alterations that cannot be easily compensated for by CAF-1-independent chromatin assembly pathways. The presence of nucleosomes and changes in nucleosome position or occupancy were analyzed using DANPOS ('dynamic analysis of nucleosome position and occupancy by sequencing') software (Chen *et al.*, 2013) (see Experimental procedures). After excluding regions of low genomic coverage (see above), DANPOS identified 563 351 nucleosomes in the reference wild-type Col genome, which corresponds to a genome-wide average density of one identified nucleosome per 210 annotated base pairs. As an earlier study reported similar nucleosome spacing of 180–190 bp (Zhang *et al.*, 2015), we conclude that most nucleosomes in *Arabidopsis* were identified in our approach.

Because CAF-1 assembles nucleosomes, we first focused on depleted nucleosomes, i.e. nucleosomes that had reduced coverage in *fas2*, which in extreme cases can be below the detection limit. There were 128 197 such depleted nucleosomes, about one-quarter of all detected nucleosomes. This demonstrates that changes in nucleosome occupancy are common in *fas2* but do not affect the entire genome. Independent experiments using H3 chromatin immunoprecipitation to evaluate nucleosome occupancy (see Experimental procedures for details) confirmed

the results of the MNase-seq: while six out of eight regions with depleted nucleosomes also had reduced occupancy in independent experiments, nucleosome occupancy was unchanged for all four control regions (Figure S1). Next, we asked whether depleted nucleosomes are enriched in certain genome regions. Because it is possible that nucleosome assembly pathways that compensate for loss of CAF-1 are related to transcription, we analyzed protein-coding gene bodies [i.e. the transcriptional units from the TSS to the transcriptional termination site (TTS)], proximal promoters (−200 to TSS), TE gene bodies and intergenic sequences. Genes that are active or inactive in corresponding leaf tissue as reported in Mozgova *et al.* (2015) (see Experimental procedures for details) were analyzed separately (Table 2). Depleted nucleosomes were detected in all analyzed genome features, but they were significantly more often found in non-transcribed regions such as proximal promoters and intergenic sequences than expected by chance (5.4% instead of 2.5% and 44.4% instead of 35.1%, respectively). Conversely, depleted nucleosomes were less often found on gene bodies than expected by chance (36.5% instead of 45.4% and 7.4% instead of 9.8% for active and inactive genes, respectively). Interestingly, active and inactive protein-coding genes as well as mostly silent TE genes all showed a similar reduction of about 0.8 in depleted nucleosomes, which strongly contrasted with the increase in depleted nucleosomes at promoters and in intergenic regions (Table 2). This result suggests that even occasional transcription can lead to re-established nucleosome occupancy. Thus, lack of CAF-1 leads to reduced nucleosome occupancy preferentially in non-transcribed genome regions. Besides the nucleosomes with reduced occupancy, about 8000 nucleosomes had increased occupancy in *fas2*. These nucleosomes were significantly enriched in bodies of active and inactive genes as well as in TE genes (Table S2). Note that the enrichment was most significant for TE genes ($P = 2.87 \times 10^{-173}$), followed by inactive genes ($P = 1.26 \times 10^{-28}$) and least significant for active genes ($P = 3.47 \times 10^{-5}$) suggesting that increased nucleosome occupancy in *fas2* is enriched at transcription

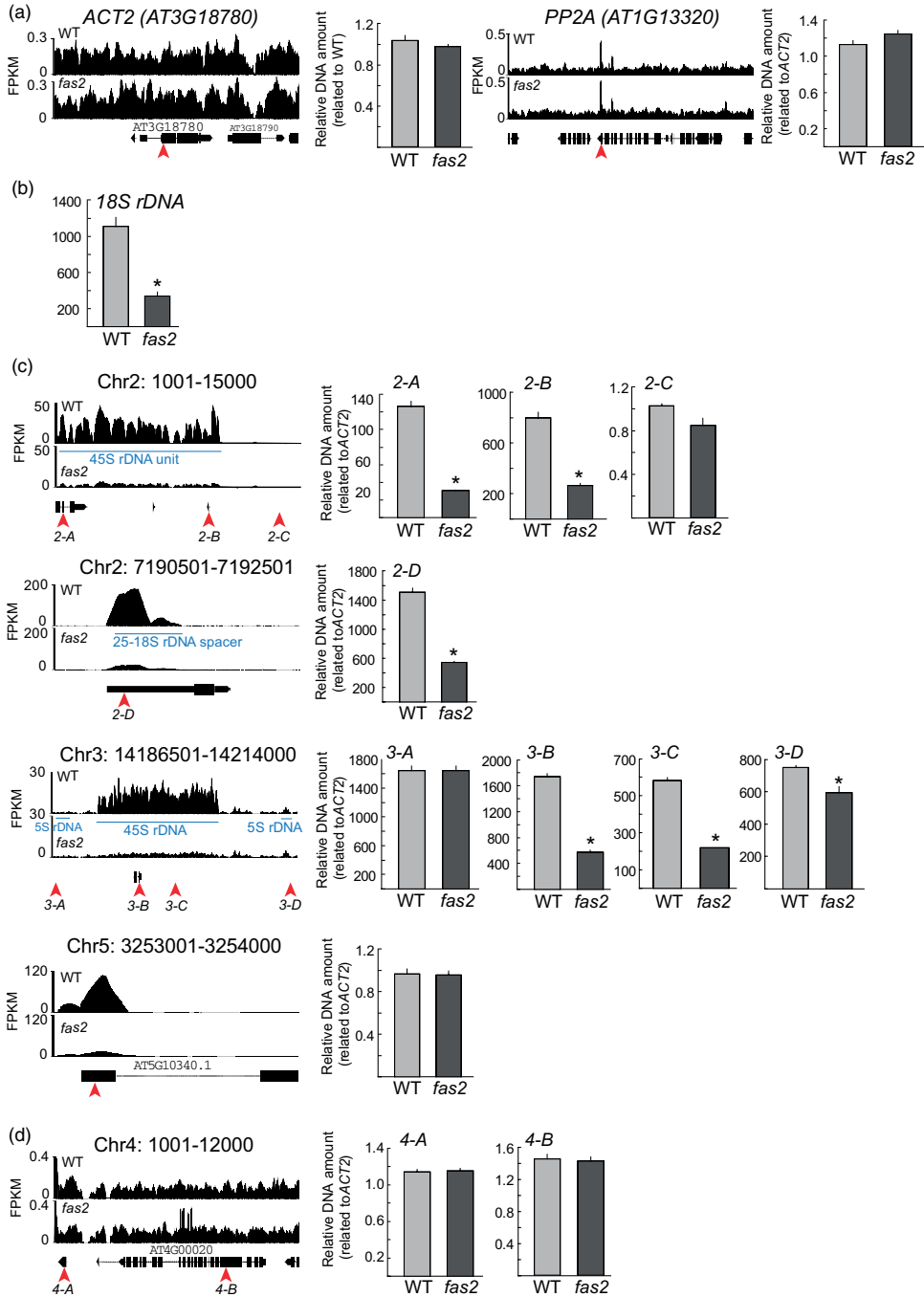


Figure 1. Confirmation of loss of genome regions.

Quantitative PCR (qPCR)-based verification of depletion of genomic DNA identified by DNA genome-wide resequencing (NGS).

(a) Single-copy regions used as non-depleted controls – *ACT2* (AT3G18780) and *PP2A* (AT1G13320). First, the amount of *ACT2* amplicon was normalized to wild-type (WT) level to confirm the comparable amount of *ACT2* DNA in *fas2* and WT. All the remaining results were normalized to *ACT2* levels.

(b) 18S rDNA-specific primers confirm the expected depletion of 45S rDNA in *fas2*.

(c) Verification of DNA depletion in the four genomic regions in *fas2* identified by NGS. Regions that were found to be repetitive relative to *ACT2* were used as queries in a BLAST (NCBI) search against the *Arabidopsis thaliana* nucleotide collection (nr/nt) and identified as 45S rDNA or 5S rDNA sequences as indicated by the blue label. Note that the depletion of the *fas2* genomic DNA region on chromosome 5 identified by NGS was not confirmed by qPCR.

(d) Genomic regions bordering the 45S rDNA cluster on chromosome 4 (NOR4) are not affected in *fas2*. Red arrowheads indicate the location of regions amplified by qPCR.

Table 2 Genome distribution of nucleosomes that are depleted in *fas2*

	No. of nucleosomes in Col	% ^a	No. of depleted nucleosomes	% ^b	FC ^c	P-value (reduced freq.)	P-value (increased freq.)
Bodies of active genes	255 563	45.4	46 800	36.5	0.8	0	1
Promoters of active genes	13 884	2.5	6957	5.4	2.17	1	0
Bodies of inactive genes	55 061	9.8	9433	7.4	0.75	2.00×10^{-255}	1
Promoters of inactive genes	4875	0.9	1603	1.3	1.39	1	1.37×10^{-59}
Bodies of TE genes	36 323	6.4	6536	5.1	0.8	7.84×10^{-117}	1
Intergenic regions	197 645	35.1	56 868	44.4	1.26	1	0

TE, transposable element.

^aPercentage of all detected nucleosomes.

^bPercentage of all depleted nucleosomes.

^cFold change of percentage of all depleted nucleosomes (i.e. observed) to percentage of all detected nucleosomes (i.e. expected).

units but is not a direct consequence of ongoing transcription. In contrast, nucleosomes with increased occupancy in *fas2* were underrepresented in non-transcribed genome regions such as promoters and intergenic sequences.

Because some repetitive genome regions are selectively lost in *Arabidopsis* CAF-1 mutants, we investigated nucleosome occupancy on telomeric and centromeric repeats on 5S rDNA and on 45S rDNA blocks (Figure S2) using a strategy proposed in Schwartz and Langst (2016). Read coverage indicated the presence of positioned nucleosomes in the wild type on telomeric and centromeric repeats but not on 5S rDNA or 45S rDNA blocks. In *fas2*, telomeres generated much less signal that did *not* indicate positioned nucleosomes, suggesting that telomeric chromatin differs substantially between the wild type and *fas2*. Similarly, centromeric repeats generated much less signal in *fas2* but the nucleosomal pattern was not lost. The analyzed 5S rDNA block contained a nucleosome-free region at the 5' end, which was lost in *fas2*, indicating that nucleosomal coverage was spread across the locus. No changes in nucleosome occupancy were obvious at the 45S rDNA. Together, nucleosome occupancy was changed on some repetitive regions in *fas2* mutant plants.

Next, we asked whether there are any particular features characteristic for nucleosomes with reduced coverage, such as the presence of DNA and histone methylation or enrichment in canonical and variant histones. Chromatin scores were calculated for DNA surrounding identified

nucleosome dyads in the wild type using 19 genome-wide epigenome datasets (see Experimental procedures for details). Although medians for all scores differed significantly between nucleosomes with reduced occupancy and other nucleosomes (Wilcoxon signed-rank test, $P < 0.05$), which is consistent with the non-random distribution of affected nucleosomes in the genome, the amplitude of the difference was always very small (Figure S3). Among the small differences, the largest effects were observed for H1 and H3K4me2, which were less abundant at wild-type nucleosomes that have reduced occupancy in *fas2*, as well as for H3.3 and H3K27me3, which were more abundant at such nucleosomes. It was unexpected to find increased H3.3 signals; however, this is consistent with the enrichment of not only nucleosomes with reduced occupancy (Table 2) but also H3.3 and H3K27me3 at many promoters (Shu *et al.*, 2014). The observation that H3.3 nucleosomes were not specifically protected from being lost in *fas2* may indicate that efficient histone replacement with H3.3 depends on undisturbed structure and composition of the chromatin substrate. Other, smaller, differences were observed for acetylated H3 (H3K9ac, H3K27ac, H3K36ac), again consistent with the presence at promoters or around the TSS. Signals for H2A.Z, H3K4me1, H3K4me3, H3K36me2, H3K36me3, H3S10p, H3S28p and DNA methylation differed least in this comparison. Together, these results argue against a specific strong epigenetic signal steering changes in nucleosome occupancy in *fas2*.

Reduced nucleosome occupancy in *fas2* preferentially in non-transcribed genome regions is consistent with the notion that CAF-1-independent nucleosome assembly is most efficient in transcribed genome regions (Gurard-Levin *et al.*, 2014). However, even in transcribed genome regions many nucleosomes were significantly depleted, suggesting that in these regions CAF-1-independent nucleosome assembly can only partially restore a normal nucleosome landscape. Earlier work had shown that transcriptional changes in this CAF-1 mutant tissue are rather limited (Mozgova *et al.*, 2015), indicating that the altered nucleosome landscape is not a direct and trivial consequence of altered gene expression. In addition, we observed maintenance of nucleosomal occupancy even in non-transcribed regions, strongly implying that CAF-1-independent nucleosome assembly mechanisms are not strictly dependent on transcription. This notion is consistent with the observation that the frequency of depleted nucleosomes did not differ considerably between active and inactive genes (18.3% and 17.1% of the detectable nucleosomes, respectively). Alternatively, prevalent genome-wide background transcription may suffice to re-establish nucleosomes even outside of annotated transcription units by a transcription-coupled nucleosome assembly process.

Next, we analyzed the distribution of depleted nucleosomes around the TSS of all protein-coding genes. Nucleosomes were binned according to their distance to the nearest TSS and *P*-values for the presence of depleted

nucleosomes were calculated per bin (Figure 2A). Depleted nucleosomes were observed more often than expected immediately upstream of the TSS, corresponding to the -1 nucleosome. Conversely, depleted nucleosomes were observed less often than expected downstream of the TSS. Notably, the same pattern of nucleosome depletion was observed for genes with unchanged (Figure 2B) or with changed expression in *fas2* (Figure 2C), suggesting that the observed change in nucleosome occupancy is not a trivial consequence of transcriptional changes in *fas2*. We note that the number of genes misregulated in *fas2* is relatively small and that the majority of genes with nucleosome depletion are not changed in expression. It was possible that nucleosome depletion affects genes independent of gene function. Alternatively, specific functional groups might be more affected than others. Gene Ontology (GO) enrichment analysis for genes with the most extreme nucleosomal changes suggested that genes were not affected randomly. Instead, genes with functions related to plant defense signaling more often had depleted nucleosomes in their promoters than other genes (Table 3). Of the 19 most significantly enriched categories ($P < 0.05$), 13 were related to plant defense. Because nucleosome depletion was most significant immediately upstream of the TSS, we focused on genes with depleted nucleosomes in the proximal promoter (-200 bp to the TSS). Again, a similar enrichment of defense-related GO categories was observed (Tables 3 and S3). Together,

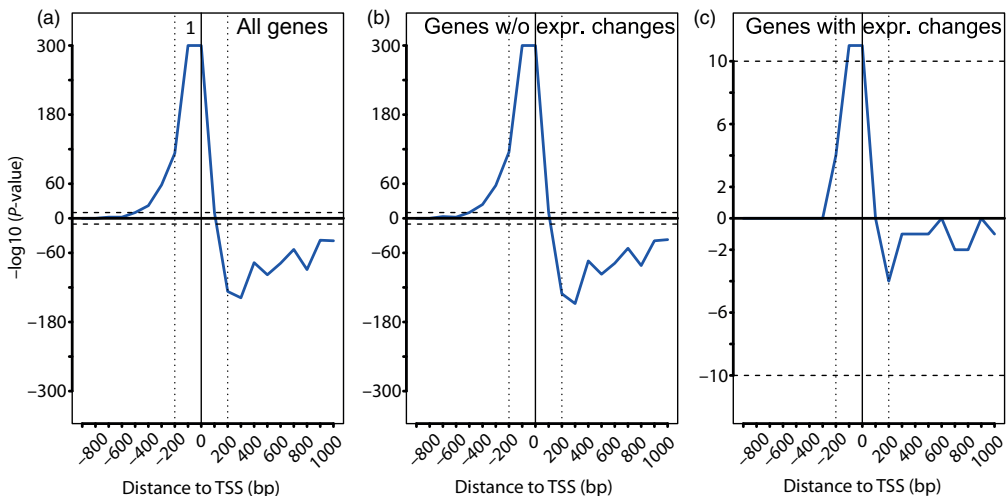


Figure 2. Localization of depleted nucleosomes.

Nucleosomes around the transcription start site (TSS) of all genes were grouped in 100-bp bins and significance of the frequency of depleted nucleosomes was calculated for each bin using hypergeometric tests followed by multiple test correction according to Bonferroni: (a) all genes, (b) genes without expression changes in *fas2*, (c) genes with expression changes in *fas2*. Horizontal dashed lines indicate $-\log_{10}(P\text{-value})$ equal to +10 and -10 , vertical solid lines indicate the TSS; vertical dashed lines indicate ± 200 bp relative to the TSS. [Colour figure can be viewed at wileyonlinelibrary.com.]

Table 3 Enriched Gene Ontology (GO) terms in genes with the 10% most depleted nucleosomes in their 1-kb promoters (–1000 TSS) or with the 20% most depleted nucleosomes their 200 bp proximal promoters (–200 TSS)

GO term	q-value
1 kb promoters^a	
Response to chitin ^b	4.90×10^{-5}
Plant-type hypersensitive response ^b	1.56×10^{-3}
Hormone-mediated signaling pathway	1.63×10^{-3}
Jasmonic acid mediated signaling pathway	4.11×10^{-3}
Negative regulation of defense response ^b	6.96×10^{-3}
Regulation of plant-type hypersensitive response ^b	7.00×10^{-3}
Salicylic acid mediated signaling pathway ^b	7.53×10^{-3}
Signal transduction	9.29×10^{-3}
Cell death	9.58×10^{-3}
Response to jasmonic acid	1.08×10^{-2}
Innate immune response ^b	1.13×10^{-2}
MAPK cascade	1.27×10^{-2}
Regulation of innate immune response ^b	1.38×10^{-2}
Response to salicylic acid ^b	1.70×10^{-2}
Defense response ^b	2.41×10^{-2}
Response to endoplasmic reticulum stress	3.06×10^{-2}
Response to fungus ^b	3.88×10^{-2}
Cellular response to abscisic acid stimulus	4.69×10^{-2}
Regulation of defense response ^b	4.69×10^{-2}
200 bp proximal promoters^a	
Response to chitin ^b	6.30×10^{-8}
Innate immune response ^b	7.70×10^{-5}
Defense response ^b	2.20×10^{-4}
Plant-type hypersensitive response ^b	5.00×10^{-4}
Respiratory burst involved in defense response ^b	9.80×10^{-4}
Hyperosmotic response	1.90×10^{-3}
Regulation of innate immune response ^b	3.60×10^{-3}
Defense response, incompatible interaction ^b	4.20×10^{-3}
Response to fungus ^b	4.40×10^{-3}
Response to other organism	5.30×10^{-3}
Regulation of plant-type hypersensitive response ^b	6.40×10^{-3}
Defense response to fungus ^b	8.90×10^{-3}
Salicylic acid mediated signaling pathway ^b	1.10×10^{-2}
Cell death	1.40×10^{-2}
Response to salicylic acid ^b	1.60×10^{-2}
Regulation of defense response ^b	3.20×10^{-2}

TSS, transcription start site; MAPK, mitogen activated protein kinase.

^aThe most depleted nucleosomes are those nucleosomes from the set of significantly depleted nucleosomes that have the largest fold change in occupancy.

^bGO terms directly related to plant defense.

nucleosome depletion in *fas2* was most pronounced in proximal promoters and preferentially affected specific functional groups of genes.

The frequent localization of depleted nucleosomes in proximal promoters suggested that CAF-1-independent nucleosome assembly mechanisms are often not efficient upstream of the TSS. The result that genes related to plant defense were particularly prone to lose nucleosomes in their promoters is consistent with earlier findings that CAF-1 deficiency causes hypersensitivity of some defense-

related genes in *Arabidopsis* (Mozgova *et al.*, 2015). It was found that the CAF-1 mutant state resembles a naturally hypersensitive (primed) state of these genes (Mozgova *et al.*, 2015). The genome-wide analysis performed here established that reduced nucleosome occupancy in promoters is not limited to the selected defense genes tested earlier but is a global effect for this class of genes. The results also reveal that nucleosomes in promoters of genes belonging to other functional classes are not as often depleted. Thus it appears that CAF-1-independent nucleosome assembly mechanisms are less efficient in promoters of defense-related genes than in other promoters. It is possible that CAF-1-independent nucleosome assembly mechanisms are repelled from promoters of defense-related genes. Alternatively, CAF-1-independent nucleosome assembly mechanisms may function at promoters of most genes but low-level defense signaling may expel nucleosomes from promoters of defense-related genes.

Because defense priming is of great importance for plant fitness, the mechanism that restricts CAF-1-independent nucleosome assembly at promoters of many defense-related genes to allow establishment and maintenance of a primed gene expression state is of considerable physiological relevance.

Nucleosome shifts in the *fas2* CAF-1 mutant are mainly in gene bodies

Nucleosomes may not only change occupancy but may also shift along the DNA. We detected 25 864 significantly shifted nucleosomes in *fas2* (see Experimental procedures and Figure S4 for details). The functional impact of a nucleosome shift depends on whether shifted nucleosomes maintain their rotational position, i.e. whether the alternative translational positions change the accessibility of bases inside nucleosomal DNA. Prevalent maintenance of rotational positions would be reflected in the preference for shifts in steps of 10–11 bp (Luger *et al.*, 1997). However, when we analyzed the distribution of shift amplitudes no evidence for any such preference could be seen (Figure S5); the 7.9% of the shifts that maintain rotational positions reflect the fraction expected by chance. Thus, nucleosome shifts in *fas2* have no obvious tendency to maintain or interrupt rotational position. In contrast to depleted nucleosomes, which were enriched in non-transcribed regions such as proximal promoters and intergenic sequences, shifted nucleosomes were enriched in gene bodies of protein-coding as well as of TE genes (Table 4). Notably, both active and inactive as well as mostly silent TE genes showed enrichment of shifted nucleosomes. In contrast, proximal promoters and intergenic sequences had fewer shifted nucleosomes than expected by chance. Next, we wondered whether, similar to nucleosome depletion in proximal promoters, nucleosome shifts in gene bodies affected specific functional groups of genes. Indeed, GO

enrichment analysis revealed that seven of the ten enriched categories (70%) were related to development (Table 5). Nucleosomes on genes can be shifted in the sense (forward) or antisense (reverse) direction. Similar numbers of nucleosomes were shifted in forward or reverse directions in *fas2* (12 847 versus 13 017). Next, we analyzed the position of shifted nucleosomes on genes in more detail. Nucleosomes were binned according to their distance to the nearest TSS and *P*-values for reduction or enrichment of shifted nucleosomes were calculated per bin (Figure 3A). As noticed above, shifted nucleosomes were enriched downstream of the TSS and reduced upstream. This was independent of the shift direction. However, the forward-shifted nucleosomes showed the strongest enrichment 100 bp downstream of the TSS corresponding to the +1 nucleosome, while reverse-shifted nucleosomes were reduced at this position and had the strongest enrichment at 400 bp downstream of the TSS. The same pattern of local enrichment of shifted nucleosomes was observed when only using genes with unchanged expression in *fas2* (Figure 3B), suggesting that, similar to the observed changes in nucleosome occupancy, the nucleosome shifts are not a trivial consequence of transcriptional changes in *fas2*.

Conclusions

This study revealed that nucleosomes that are depleted in an *Arabidopsis* CAF-1 mutant are enriched in non-transcribed regions. This result is consistent with the notion that CAF-1-independent nucleosome assembly is most efficient in transcribed genome regions. Because the analyzed tissue did not contain replicating nuclei, additional chromatin changes may exist in CAF-1 mutants transiently during or after replication but were masked by post-replicative CAF-1-independent nucleosome assembly. The efficiency of CAF-1-independent nucleosome assembly at promoters differs for different functional groups of genes, indicating the action of a not well understood mechanism of gene regulation. The mechanism limiting CAF-1-independent nucleosome assembly at promoters of many defense-related genes is important for plant fitness as it allows establishment and maintenance of a primed gene expression state that contributes to an efficient defense against pathogens.

EXPERIMENTAL PROCEDURES

Plant material and growth conditions

The *Arabidopsis thaliana* wild-type and T-DNA insertion lines were in the Columbia-0 (Col) background; *fas2-4* (SALK_033228) was described previously (Exner et al., 2006). Because CAF-1 mutants suffer from a progressive loss of tandem repeats and vary between generations (Mozgova et al., 2010), all adult-plant experiments were performed on *fas2-4* generation 6 (G6) plants. Seeds were sown on 0.5 × basal salts Murashige and Skoog (MS) medium (Duchefa, <https://www.duchefa-biochemie.com/>), stratified at

4°C for 48 h, and allowed to germinate for 10 days in growth chambers under short-day (8-h light, 100 μmol m⁻² sec⁻¹, 22°C) conditions. Plantlets were planted in soil and grown in growth chambers under the same conditions. For chromatin extraction, leaf number 6 was harvested after 38 days at zeitgeber time (ZT) = 7 h. At this time, leaf number 6 is fully expanded and all cell division and endoreduplication has ceased (Baerenfaller et al., 2012). Note that this condition corresponds to standard conditions defined in the AGRON-OMICS project that have been used in several studies, facilitating data comparison (Baerenfaller et al., 2012, 2015; Derkacheva et al., 2013; Shu et al., 2013, 2014; Mozgova et al., 2015).

Genome resequencing

For genome resequencing, generation 6 (G6) *fas2-4* and corresponding segregated wild-type control plants were used. Ten-day old seedlings were harvested and DNA was isolated using a MagJET Plant Genomic DNA Kit (Thermo Fisher Scientific, <https://www.thermofisher.com/>) according to the manufacturer's instructions. Genomic DNA sequencing libraries were generated using a TruSeq DNA LT Library Preparation kit (Illumina Inc., <https://www.illumina.com/>) and sequenced using the 100-bp paired-end mode on an Illumina HiSeq 2000 system (Illumina Inc.). Reads with a total sequencing quality of less than 30 Phred or shorter than 10 nucleotides (nt) were discarded. For each library, 25 million clean paired-end reads were mapped to the reference genome TAIR10 using bwa with default parameters (Li and Durbin, 2009b). The mapping efficiency was 94% and 87% for the Col and *fas2* mutant libraries, respectively. Duplicated paired mappings were removed using samtools rmdup (Li et al., 2009a). The annotated nuclear genome was divided into 1 kb bins and the number of mapped reads falling into each bin were counted. Bins containing fewer reads in Col than the tenth lower quantile of all bins (67 reads per 1 kb bin) were not considered for further analysis. A bin-wise one-sided binomial test was used to test for reduced coverage in *fas2*. Multiple testing correction was done according to (Storey and Tibshirani, 2003). Bins were considered to have reduced coverage in *fas2* if $q < 0.05$ and \log_2 fold change (FC) < -1.5. To account for different numbers of mapped reads in the wild type and *fas2*, coverage was expressed in FKPM (fragments per 1 kilobase and 1 million mapped fragments).

MNase-seq

Leaf material was cross-linked using 1% formaldehyde under vacuum for 15 min. Cross-linking was quenched using an excess amount of glycine (0.125 M) and treatment with vacuum for another 5 min. The leaves were rinsed with water, blot-dried on filter paper and ground in liquid nitrogen. To extract chromatin, 100 mg of ground frozen leaf material was treated in Nuclei Extraction Buffer [NEB; 20 mM piperazine-*N,N*-bis(2-ethanesulfonic acid)-KOH pH 7.6, 1 M hexylene glycol, 10 mM MgCl₂, 0.1 mM EGTA, 15 mM NaCl, 60 mM KCl, 0.5% Triton-X, 5 mM β-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche, <http://www.roche.com/>)] for 15 min at 4°C. The homogenate was filtered through a CellTrics nylon filter (50 μm) (Partec, <https://www.sysmex-partec.com/>) and pellets were collected by centrifugation for 10 min at 1500 *g* at 4°C. Pellets were washed once in MNase buffer [50 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (TRIS)-HCl pH 8, 10 mM NaCl, 5 mM CaCl₂ and EDTA-free protease inhibitor cocktail (Roche)] and were re-suspended in 100 μl of MNase buffer. An aliquot of 20 μl was set aside as an input control, and 20 μl of MNase

Table 4 Genome distribution of nucleosomes that are shifted in *fas2*

	No. of nucleosomes in Col	% ^a	No. of shifted nucleosomes	% ^b	FC ^c	P-value (reduced freq.)	P-value (increased freq.)
Bodies of active genes	255 563	45.4	12 628	48.8	1.08	1	1.48×10^{-30}
Promoters of active genes	13 884	2.5	303	1.2	0.47	2.72×10^{-52}	1
Bodies of inactive genes	55 061	9.8	2945	11.4	1.16	1	9.84×10^{-19}
Promoters of inactive genes	4875	0.9	148	0.6	0.64	2.04×10^{-8}	1
Bodies of TE genes	36 323	6.4	2469	9.5	1.49	1	1.92×10^{-85}
Intergenic regions	197 645	35.1	7371	28.5	0.81	4.55×10^{-118}	1

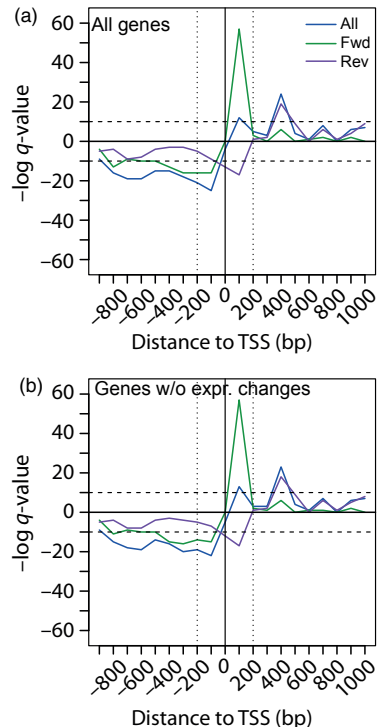
TE, transposable element.

^aPercentage of all detected nucleosomes.^bPercentage of all shifted nucleosomes.^cFold change of percentage of all shifted nucleosomes (i.e. observed) to percentage of all detected nucleosomes (i.e. expected).**Table 5** Enriched Gene Ontology (GO) terms in genes with the 10% most shifted nucleosomes in their gene bodies

GO term	q-value
Transmembrane transport	2.42×10^{-3}
Reproductive structure development ^a	2.81×10^{-3}
Multicellular organismal development ^a	2.91×10^{-3}
Shoot system development ^a	3.41×10^{-3}
Tissue development ^a	1.70×10^{-2}
Flower development ^a	2.60×10^{-2}
Anatomical structure morphogenesis ^a	2.76×10^{-2}
Auxin efflux	3.65×10^{-2}
Meristem maintenance ^a	3.83×10^{-2}
Regulation of cell proliferation	3.84×10^{-2}

^aGO terms directly related to plant development.

buffer and 1.3 μ l of RNase A (30 μ g μ l⁻¹; Sigma-Aldrich, <http://www.sigmaaldrich.com/>) were added to the rest of the suspension. This mixture was digested with MNase (New England Bio-Labs, <https://www.neb.com/>; final concentration 0.2 U μ l⁻¹) for 8 min in MNase buffer. The reaction was stopped with 10 mM EDTA. After centrifugation for 5 min at 1500 *g* at 4°C the supernatant was collected as supernatant 1. The pellet was re-suspended in S2 buffer [1 mM TRIS-HCl pH 8, 0.2 mM EDTA and EDTA-free protease inhibitor cocktail (Roche)] and treated for 30 min on ice. The suspension was centrifuged again for 5 min at 1500 *g* at 4°C and the supernatant was collected as supernatant 2. After pooling supernatants 1 and 2, formaldehyde cross-links were reversed by treatment at 65°C overnight with elevated salt concentration (1 M NaCl). The mixture was treated with de-proteinization solution [0.02 M EDTA, 0.1 M TRIS-HCl pH 6.5, 1 μ l Protease K (Sigma)] at 45°C for 3 h. Nucleosome DNA was recovered as above. Digested DNA was resolved on 1% agarose gels. Gel slabs containing the DNA fraction 150 \pm 50 bp were sliced and DNA was recovered using a GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences, <http://www.gelifesciences.com>) according to the manufacturer's protocol. Sequencing libraries were constructed using an Ovation Ultralow Library System (NuGEN, <http://www.nugen.com/>), including eight different barcodes as described in the manufacturer's protocol. Sequencing was performed on an Illumina HiSeq 2000 System in 100-bp paired-end mode using v.3 sequencing chemistry at the Science for Life Laboratories in Uppsala, Sweden (<https://www.scilifelab.se/>).

**Figure 3.** Localization of shifted nucleosomes.

Nucleosomes around the transcription start site (TSS) of all genes were grouped in 100-bp bins and the significance of enrichment of shifted nucleosomes was calculated for each bin using hypergeometric tests followed by multiple test correction according to Bonferroni. Tests were performed for all shifted nucleosomes (blue lines), nucleosomes shifted forward (green lines) or in reverse (purple lines) relative to the direction of transcription: (a) all genes and (b) genes without expression changes in *fas2*. Horizontal dashed lines indicate $-\log_{10}(P\text{-value})$ equal to +10 and -10; vertical solid lines indicate the TSS; vertical dashed lines indicate ± 200 bp relative to the TSS.

Bioinformatics analysis

FastQC v.0.10.1 was used to check read quality. Libraries of 30.3 and 39.4 million paired reads for Col and *fas2*, respectively, were mapped with bowtie2 against the TAIR10 genome with efficiencies of 85% using the local option. Mapped MNase-seq DNA samples were analyzed with DANPOS (Chen *et al.*, 2013), which extracted the position of nucleosome dyads and performed statistical tests to identify significantly changed nucleosomes. Briefly, DANPOS takes as input two paired-end mapped bam files, removes read duplicates, calculates the position of the nucleosome adjusting the forward and reverse reads to the middle of their positions and calculates nucleosome occupancy as local read coverage. Then DANPOS compares the two samples by first normalizing samples using quantile normalization and secondly performing differential signal calculation with a Poisson test. DANPOS identifies the position of the nucleosome dyad. A nucleosome is considered as shifted when the position of the dyad in sample 1 is 50–90 nt away from the position in sample 2. Small differences in dyad position (<50 nt) are considered noise rather than a shift; large differences in dyad position (>90 nt) are considered to reflect different nucleosomes rather than shifts. Figure S4 illustrates the method of assigning nucleosome shifts. For more details on algorithms and tests used by DANPOS see Chen *et al.* (2013). Further analysis was performed with custom scripts in R v.3.1.2. A GO analysis was performed using the GOstats R package. The hypergeometric test with Bonferroni correction was used with a filter requiring three genes with a manually assigned annotation as minimum category population. Transcriptome data are from (Mozgova *et al.*, 2015) based on Affymetrix tiling arrays with a signal of 5 as threshold to differentiate active from inactive genes (Rehrauer *et al.*, 2010). That study identified 307 upregulated and 65 downregulated genes in fully grown leaves of *fas2-4*, which correspond to the material used here. Thus, the vast majority of genes are not differentially expressed in this material. For the epigenetic profiling of depleted and non-depleted nucleosomes, scores were calculated by averaging the Z-score coverage values ± 50 bp around the nucleosome dyad detected by DANPOS in wild-type conditions. Chromatin profiles were taken from the literature: H1.1 and H1.2 from Rutowicz *et al.* (2015); H3K27ac, H3K27me3, H3K4me3, H3K9ac, H3K9me2 and H3S10p Baerenfaller *et al.* (2016); H3.3 Shu *et al.* (2014); H2A.Z Coleman-Derr and Zilberman, 2012); H3K4me1 Zhang *et al.*, 2009); H3K4me2 and H3K36me2 Luo *et al.*, 2013); H3K36ac and H3K36me3 Mahrez *et al.* (2016); DNA inaccessibility Shu *et al.*, 2012) and DNA methylation Stroud *et al.* (2013) H1.1 and H1.2: 3-week-old plants; H3K27ac, H3K27me3, H3K4me3, H3K9ac, H3K9me2, H3S10p, H3K36ac, H3K36me3, H3.3 and DNA inaccessibility: leaf number 6, 5-week-old plants; H2A.Z: 4 weeks' post-germination roots; H3K4me and DNA methylation: 3-week-old plants; H3K4me2 and H3K36me2: 2-week-old plants.

Experimental confirmation of loss of genome regions

Genomic DNA (gDNA) was isolated from 100 mg (fresh weight) 10-day-old seedlings using the MagJET Plant Genomic DNA Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Concentration of the purified RNA-free DNA was quantified using the Qubit double-stranded DNA (dsDNA) BR (broad range) Assay Kit (Thermo Fisher Scientific) and adjusted to 10 ng μl^{-1} in all samples. Then 0.02–20 ng of gDNA were used in a 20- μl quantitative PCR reaction using 5 \times HOT FIREPol® EvaGreen® quantitative PCR (qPCR) Supermix (Solis Biodyne,

<https://www.sbd.ee/>) and primers specified in Table S4. Quantification was done using a MyiQTM Single Color Real Time PCR detection system (Bio-Rad, <http://www.bio-rad.com/>), the results were normalized to *ACT2* (*At3G18780*) and to the wild-type seedlings. The experiment was performed in three biological replicates.

Chromatin immunoprecipitation (ChIP)

Fully expanded leaves from 5-week-old Col and *fas2* plants grown under long-day conditions were collected and cross-linked using 1% formaldehyde for 10 min followed by quenching by 0.125 M glycine for 5 min under vacuum. One hundred milligrams of ground material was used to extract crude nuclei as described above for MNase-sequencing. Nuclei were washed once using 200 μl of MNase buffer and collected by centrifugation (1500 g, 10 min, 4°C). Extracted nuclei were re-suspended in 50 μl of MNase buffer and 10 μl of non-digested sample was taken as an input control. Five units of MNase (New England BioLabs) were added and the suspension was incubated for 6 min at 37°C. The reaction was stopped by adding EDTA (final 10 mM) and SDS (final 1%). One cycle of 30 sec sonication was applied (Bioruptor sonicator, Diagenode, <https://www.diagenode.com/>), the suspension was diluted to 500 μl using ChIP buffer [16.7 mM TRIS-HCl, pH 8.0, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 1 \times cOmplete EDTA-free proteinase inhibitor (Roche)], and the chromatin was cleared by centrifugation at 4500 g (5 min, 4°C). Immunoprecipitation was performed as described Mozgova *et al.* (2015). For each immunoprecipitation, 200 μl of cleared chromatin was incubated with 1.5 μg of anti-histone H3 antibody (Merc Millipore, <http://www.merckmillipore.com> cat. no. 07-690) or IgG (Sigma, cat. no. I5006) as a control. Immunoprecipitated and input DNA was de-crosslinked, treated with Proteinase K (Sigma), extracted using phenol-chloroform and quantified in technical duplicates by qPCR on the RotorGene RG-3000 system (Corbett Research, <https://www.qiagen.com/>) using 5 \times HOT FIREPol Eva Green qPCR Mix Plus (ROX) (Solis Biodyne) and primers specified in Table S5. The amount of DNA was related to the input control before MNase digestion. Genomic regions representing nucleosomes with reduced occupancy in *fas2*, as identified by the MNase-seq approach, were selected for confirmation by MNase-ChIP-qPCR. Only single-copy regions where the closest neighboring gene did not differ in expression level in *fas2* (Mozgova *et al.*, 2015) were considered. Regions with no changes in nucleosome occupancy or position were taken as controls. Primers were designed to amplify regions spanning approximately 50 bp upstream and 50 bp downstream of the wild-type nucleosome dyad axis position. ChIP was performed in triplicates using leaves of independent plants; biological replicates were averaged and a lower anti-H3 DNA recovery in *fas2* than in the wild type was tested using a one-sided paired Student's *t*-test.

ACCESSION NUMBER

Data are available at GEO (accession number GSE87421).

ACKNOWLEDGEMENT

This work was supported by grants from the Swiss National Science Foundation SNF, the Swedish Science Foundation VR and the Knut-and-Alice-Wallenberg Foundation.

CONFLICT OF INTEREST

The authors confirm that they have no conflicts of interest to declare.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Sequence coverage in selected repeat regions.

Table S2. Genome distribution of nucleosomes with increased occupancy in *fas2*.

Table S3. Enriched Gene Ontology terms in genes with depleted nucleosomes in their proximal promoters (–200 transcription start site).

Table S4. Primers used for verification of genome resequencing using quantitative PCR.

Table S5. Primers used for verification of micrococcal nuclease sequencing using quantitative PCR.

Figure S1. Validation of micrococcal nuclease sequencing results.

Figure S2. Repetitive regions did not undergo selective loss of nucleosomes in *fas2*.

Figure S3. Epigenetic profile of depleted and non-depleted nucleosomes.

Figure S4. Scheme of the method for assigning nucleosome shifts.

Figure S5. The magnitude of the shift of displaced nucleosomes in *fas2* does not follow a periodic trend.

REFERENCES

- Abe, M., Kuroshita, H., Umeda, M., Itoh, J. and Nagato, Y. (2008) The rice flattened shoot meristem, encoding CAF-1 p150 subunit, is required for meristem maintenance by regulating the cell-cycle period. *Dev. Biol.* **319**, 384–393.
- Baerenfaller, K., Massonnet, C., Walsh, S. *et al.* (2012) Systems-based analysis of *Arabidopsis* leaf growth reveals adaptation to water deficit. *Mol. Syst. Biol.* **8**, 606.
- Baerenfaller, K., Massonnet, C., Hennig, L., Russenberger, D., Sulpice, R., Walsh, S., Stitt, M., Granier, C. and Grissem, W. (2015) A long photoperiod relaxes energy management in *Arabidopsis* leaf six. *Curr. Plant Biol.* **2**, 34–45.
- Baerenfaller, K., Shu, H., Hirsch-Hoffmann, M., Futterer, J., Opitz, L., Rehrauer, H., Hennig, L. and Grissem, W. (2016) Diurnal changes in the histone H3 signature H3K9ac/H3K27ac/H3S28p are associated with diurnal gene expression in *Arabidopsis*. *Plant, Cell Environ.* **39**, 2557–2569.
- Chen, Z., Tan, J.L., Ingouff, M., Sundaresan, V. and Berger, F. (2008) Chromatin assembly factor 1 regulates the cell cycle but not cell fate during male gametogenesis in *Arabidopsis thaliana*. *Development*, **135**, 65–73.
- Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X. and Li, W. (2013) DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351.
- Coleman-Derr, D. and Zilberman, D. (2012) Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet.* **8**, e1002988. LID-<https://doi.org/10.1371/journal.pgen.1002988>.
- Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgova, I., Mahrez, W., Nanni, P., Bischof, S., Grissem, W. and Hennig, L. (2013) *Arabidopsis* MSI1 connects LHP1 to PRC2 complexes. *EMBO J.* **32**, 2073–2085.
- Duc, C., Benoit, M., Le Goff, S., Simon, L., Poulet, A., Cotterell, S., Tatout, C. and Probst, A.V. (2015) The histone chaperone complex HIR maintains nucleosome occupancy and counterbalances impaired histone deposition in CAF-1 complex mutants. *Plant J.* **81**, 707–722.
- Endo, M., Ishikawa, Y., Osakabe, K. *et al.* (2006) Increased frequency of homologous recombination and T-DNA integration in *Arabidopsis* CAF-1 mutants. *EMBO J.* **25**, 5579–5590.
- Exner, V., Taranto, P., Schönrock, N., Grissem, W. and Hennig, L. (2006) Chromatin assembly factor CAF-1 is required for cellular differentiation during plant development. *Development*, **133**, 4163–4172.
- Exner, V., Grissem, W. and Hennig, L. (2008) Control of trichome branching by Chromatin Assembly Factor-1. *BMC Plant Biol.* **8**, 54.
- Guittón, A.E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.E., Grossniklaus, U. and Berger, F. (2004) Identification of new members of FERTILISATION INDEPENDENT SEED Polycomb group pathway involved in the control of seed development in *Arabidopsis thaliana*. *Development*, **131**, 2971–2981.
- Guarard-Levin, Z.A., Quivy, J.P. and Almouzni, G. (2014) Histone chaperones: assisting histone traffic and nucleosome dynamics. *Annu. Rev. Biochem.* **83**, 487–517.
- Havlova, K., Dvorackova, M., Peiro, R., Abia, D., Mozgova, I., Vanscova, L., Gutierrez, C. and Fajkus, J. (2016) Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. *Plant Mol. Biol.* **92**, 457–471.
- Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M. and Henikoff, S. (2011) Epigenome characterization at single base-pair resolution. *Proc. Natl Acad. Sci. USA*, **108**, 18318–18323.
- Hennig, L., Bouveret, R. and Grissem, W. (2005) MSI1-like proteins: an escort service for chromatin assembly and remodeling complexes. *Trends Cell Biol.* **15**, 295–302.
- Hoek, M. and Stillman, B. (2003) Chromatin Assembly Factor-1 is essential and couples chromatin assembly to DNA replication *in vivo*. *Proc. Natl Acad. Sci. USA*, **100**, 12183–12188.
- Houlard, M., Berlivet, S., Probst, A.V., Quivy, J.P., Hery, P., Almouzni, G. and Gerard, M. (2006) CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. *PLoS Genet.* **2**, 1686–1696.
- Jin, C. and Felsenfeld, G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.* **21**, 1519–1529.
- Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark ‘nucleosome-free regions’ of active promoters and other regulatory regions. *Nat. Genet.* **41**, 941–945.
- Kaufman, P.D., Kobayashi, R., Kessler, N. and Stillman, B. (1995) The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell*, **81**, 1105–1114.
- Kaufman, P.D., Kobayashi, R. and Stillman, B. (1997) Ultraviolet radiation sensitivity and reduction of telomeric silencing in *Saccharomyces cerevisiae* cells lacking Chromatin Assembly Factor-1. *Genes Dev.* **11**, 345–357.
- Kaufman, P.D., Cohen, J.L. and Osley, M.A. (1998) Hir proteins are required for position-dependent gene silencing in *Saccharomyces cerevisiae* in the absence of chromatin assembly factor I. *Mol. Cell. Biol.* **18**, 4793–4806.
- Kaya, H., Shibahara, K., Taoka, K., Iwabuchi, M., Stillman, B. and Araki, T. (2001) FASCIATA genes for Chromatin Assembly Factor-1 in *Arabidopsis* maintain the cellular organization of apical meristems. *Cell*, **104**, 131–142.
- Kirik, A., Pecinka, A., Wendeler, E. and Reiss, B. (2006) The Chromatin Assembly Factor subunit FASCIATA1 is involved in homologous recombination in plants. *Plant Cell*, **18**, 2431–2442.
- Köhler, C., Hennig, L., Bouveret, R., Gheysels, J., Grossniklaus, U. and Grissem, W. (2003) *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO J.* **22**, 4804–4814.
- Leyser, H.M. and Furrer, I.J. (1992) Characterisation of three shoot apical meristem mutants of *Arabidopsis thaliana*. *Development*, **116**, 397–403.
- Li, H. and Durbin, R. (2009b) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009a) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
- Luo, C., Sidote, D.J., Zhang, Y., Kerstetter, R.A., Michael, T.P. and Lam, E. (2013) Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. *Plant J.* **73**, 77–90.
- Mahrez, W., Trejo Arellano, M.S., Moreno-Romero, J., Nakamura, M., Shu, H., Nanni, P., Köhler, C., Grissem, W. and Hennig, L. (2016) H3K36ac is an evolutionary conserved plant histone modification that marks active genes. *Plant Physiol.* **170**, 1566–1577.

- Mozgova, I., Mokros, P. and Fajkus, J. (2010) Dysfunction of Chromatin Assembly Factor-1 induces shortening of telomeres and loss of 45S rDNA in *Arabidopsis thaliana*. *Plant Cell*, **22**, 2768–2780.
- Mozgova, I., Wildhaber, T., Liu, Q., Abou-Mansour, E., L'Haridon, F., Métraux, J., Grissem, W., Hofius, D. and Hennig, L. (2015) Chromatin assembly factor CAF-1 represses priming of plant defence response genes. *Nat. Plants*, **1**, 15127.
- Muchova, V., Amiard, S., Mozgova, I., Dvorackova, M., Gallego, M.E., White, C. and Fajkus, J. (2015) Homology-dependent repair is involved in 45S rDNA loss in plant CAF-1 mutants. *Plant J.* **81**, 198–209.
- Murata, M., Ogura, Y. and Motoyoshi, F. (1994) Centromeric repetitive sequences in *Arabidopsis thaliana*. *Jpn. J. Genet.* **69**, 361–370.
- Nabatiyan, A. and Krude, T. (2004) Silencing of Chromatin Assembly Factor-1 in human cells leads to cell death and loss of chromatin assembly during DNA synthesis. *Mol. Cell. Biol.* **24**, 2853–2862.
- Ono, T., Kaya, H., Takeda, S., Abe, M., Ogawa, Y., Kato, M., Kakutani, T., Scheid, O.M., Araki, T. and Shibahara, K. (2006) Chromatin Assembly Factor-1 ensures the stable maintenance of silent chromatin states in *Arabidopsis*. *Genes Cells*, **11**, 153–162.
- Pavlistova, V., Dvorackova, M., Jez, M., Mozgova, I., Mokros, P. and Fajkus, J. (2016) Phenotypic reversion in fas mutants of *Arabidopsis thaliana* by reintroduction of FAS genes: variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes. *Plant J.* **88**, 411–424.
- Ramirez-Parra, E. and Gutierrez, C. (2007a) E2F regulates *FASCIATA1*, a chromatin assembly gene whose loss switches on the endocycle and activates gene expression by changing the epigenetic status. *Plant Physiol.* **144**, 105–120.
- Ramirez-Parra, E. and Gutierrez, C. (2007b) The many faces of Chromatin Assembly Factor 1. *Trends Plant Sci.* **12**, 570–576.
- Ray-Gallet, D., Woolfe, A., Vassias, I. et al. (2011) Dynamics of histone H3 deposition *in vivo* reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol. Cell*, **44**, 928–941.
- Rehrauer, H., Aquino, C., Grissem, W. et al. (2010) AGRONOMICS1: a new resource for *Arabidopsis* transcriptome profiling. *Plant Physiol.* **152**, 487–499.
- Reinholz, E. (1966) Radiation induced mutants showing changed inflorescence characteristics. *Arab. Inf. Serv.* **3**, 19–20.
- Rutowicz, K., Puzio, M., Halibart-Puzio, J. et al. (2015) A specialized Histone H1 Variant is required for adaptive responses to complex abiotic stress and related DNA methylation in *Arabidopsis*. *Plant Physiol.* **169**, 2080–2101.
- Schönrock, N., Exner, V., Probst, A., Grissem, W. and Hennig, L. (2006) Functional genomic analysis of CAF-1 mutants in *Arabidopsis thaliana*. *J. Biol. Chem.* **281**, 9560–9568.
- Schwartz, U. and Langst, G. (2016) Bioinformatic Analysis of ChIP-seq Data on the Repetitive Ribosomal RNA Gene. *Methods Mol. Biol.* **1455**, 225–230.
- Shibahara, K. and Stillman, B. (1999) Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell*, **96**, 575–585.
- Shu, H., Wildhaber, T., Siretskiy, A., Grissem, W. and Hennig, L. (2012) Distinct modes of DNA accessibility in plant chromatin. *Nat. Commun.* **3**, 1281.
- Shu, H., Grissem, W. and Hennig, L. (2013) Measuring *Arabidopsis* chromatin accessibility using DNase I-polymerase chain reaction and DNase I-chip assays. *Plant Physiol.* **162**, 1794–1801.
- Shu, H., Nakamura, M., Siretskiy, A., Borghi, L., Moraes, I., Wildhaber, T., Grissem, W. and Hennig, L. (2014) *Arabidopsis* replacement histone variant H3.3 occupies promoters of regulated genes. *Genome Biol.* **15**, R62.
- Smith, S. and Stillman, B. (1989) Purification and characterization of CAF-1, a human cell factor required for chromatin assembly during DNA replication *in vitro*. *Cell*, **58**, 15–25.
- Song, Y., He, F., Xie, G. et al. (2007) CAF-1 is essential for *Drosophila* development and involved in the maintenance of epigenetic memory. *Dev. Biol.* **311**, 213–222.
- Stillman, B. (1986) Chromatin assembly during sv40 DNA replication *in vitro*. *Cell*, **45**, 555–565.
- Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genome-wide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
- Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V. and Jacobsen, S.E. (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell*, **152**, 352–364.
- Tagami, H., Ray-Gallet, D., Almouzni, G. and Nakatani, Y. (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell*, **116**, 51–61.
- Talbert, P.B. and Henikoff, S. (2010) Histone variants—ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**, 264–275.
- Winkler, D.D., Zhou, H., Dar, M.A., Zhang, Z. and Luger, K. (2012) Yeast CAF-1 assembles histone (H3-H4)2 tetramers prior to DNA deposition. *Nucleic Acids Res.* **40**, 10139–10149.
- Xi, Y., Yao, J., Chen, R., Li, W. and He, X. (2011) Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome Res.* **21**, 718–724.
- Yu, Z., Liu, J., Deng, W.M. and Jiao, R. (2015) Histone chaperone CAF-1: essential roles in multi-cellular organism development. *Cell. Mol. Life Sci.* **72**, 327–337.
- Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M. and Jacobsen, S.E. (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **10**, R62.
- Zhang, T., Zhang, W. and Jiang, J. (2015) Genome-wide nucleosome occupancy and positioning and their impact on gene expression and evolution in plants. *Plant Physiol.* **168**, 1406–1416.

H3K36ac Is an Evolutionary Conserved Plant Histone Modification That Marks Active Genes^{1[OPEN]}

Walid Mahrez, Minerva Susana Trejo Arellano, Jordi Moreno-Romero, Miyuki Nakamura, Huan Shu, Paolo Nanni, Claudia Köhler, Wilhelm Gruissem, and Lars Hennig*

Department of Plant Biology and Linnean Center for Plant Biology, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden (W.M., M.S.T.A., J.M.-R., M.N., C.K., L.H.); Department of Biology and Zurich-Basel Plant Science Center, ETH Zurich, CH-8092 Zurich, Switzerland (W.M., H.S., W.G.); and Functional Genomics Center Zurich, University of Zurich/ETH Zurich, CH-8057 Zurich, Switzerland (P.N.)

ORCID IDs: 0000-0003-3771-5221 (W.M.); 0000-0002-7352-1507 (J.M.-R.); 0000-0002-8153-7293 (M.N.); 0000-0001-8429-3557 (P.N.); 0000-0002-2619-4857 (C.K.); 0000-0002-6645-1862 (L.H.)

In eukaryotic cells, histones are subject to a large number of posttranslational modifications whose sequential or combinatorial action affects chromatin structure and genome function. We identified acetylation at Lys-36 in histone H3 (H3K36ac) as a new chromatin modification in plants. The H3K36ac modification is evolutionarily conserved in seed plants, including the gymnosperm Norway spruce (*Picea abies*) and the angiosperms rice (*Oryza sativa*), tobacco (*Nicotiana tabacum*), and Arabidopsis (*Arabidopsis thaliana*). In Arabidopsis, H3K36ac is highly enriched in euchromatin but not in heterochromatin. Genome-wide chromatin immunoprecipitation sequencing experiments revealed that H3K36ac peaks at the 5' end of genes, mainly on the two nucleosomes immediately distal to the transcription start site, independently of gene length. H3K36ac overlaps with H3K4me3 and the H2A.Z histone variant. The histone acetyl transferase GCN5 and the histone deacetylase HDA19 are required for H3K36ac homeostasis. H3K36ac and H3K36me3 show negative crosstalk, which is mediated by GCN5 and the histone methyl transferase SDG8. Although H3K36ac is associated with gene activity, we did not find a linear relationship between H3K36ac and transcript levels, suggesting that H3K36ac is a binary indicator of transcription.

Nuclear DNA of eukaryotes is packed in chromatin, the basic unit of which is the nucleosome, consisting of 147 bp of DNA wrapped around a protein octamer formed by two copies each of four histones (H2A, H2B, H3, and H4). Chromatin structure and composition are dynamically regulated both temporally and spatially within the nucleus. Covalent modifications of histones play an important role in the regulation of gene expression by changing chromatin structure (Strahl and Allis, 2000; Jenuwein and Allis, 2001). Many histone modifications have been identified, including acetylation, methylation, phosphorylation, and ubiquitinylation (Cannon, 1955; Zhang et al., 2007). It has been proposed that histone modifications act in a combinatorial manner to establish specific chromatin states that regulate gene

expression (Strahl and Allis, 2000; Jenuwein and Allis, 2001). It has been argued, however, that chromatin-dependent gene activity is only a consequence of the cumulative effect of histone modifications rather than the interpretation of an alphabet (Henikoff, 2005). Using high-throughput methods with improved sensitivity, specificity, and resolution such as mass spectrometry and chromatin immunoprecipitation (ChIP) coupled to sequencing (ChIP-seq), the genomic localization, combinatorial patterns, and enrichment of histone modifications can be resolved at the level of individual nucleosomes (Liu et al., 2005; Kharchenko et al., 2011).

Histone acetylation is a dynamic and reversible modification, in which an acetyl group is transferred from acetyl-CoA to histone Lys residues. Lys acetylation neutralizes positive charges of the histone tails, thus decreasing their affinity for negatively charged DNA (Hong et al., 1993). Acetylated lysines can be recognized by histone code readers such as Bromo domain proteins (Sanchez et al., 2014), which then may promote the recruitment of additional chromatin remodeler or modifier complexes to relax chromatin and facilitate the binding or activation of RNA polymerases (Bannister and Kouzarides, 2011). Histone Lys acetylation is often associated with transcription and has important roles in numerous developmental and biological processes, including the regulation of cell cycle, flowering time, response to environmental conditions, and hormone signaling (Chen and Tian, 2007; Luo et al., 2012).

¹ This work was supported by grants from the Swiss National Science Foundation, the Swedish Science Foundation VR, and the Knut-and-Alice-Wallenberg Foundation.

* Address correspondence to lars.hennig@slu.se.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Lars Hennig (lars.hennig@slu.se).

W.M., J.M.-R., M.N., and H.S. performed the experiments; W.M., M.S.T.A., H.S., P.N., and L.H., analyzed data; W.M., C.K., and L.H. planned the experiments; W.M., M.S.T.A., C.K., W.G., and L.H. wrote the manuscript.

^[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.15.01744

Several N-terminal Lys residues of histone H3 (K9, K14, K18, K23, and K27) and H4 (K5, K8, K12, K16, and K20) are known to be acetylated in *Arabidopsis thaliana*; Berr et al., 2011). Histone acetylation is dynamically regulated with a usual half-life of 2 to 3 min in human cell lines that rarely exceeds 30 to 40 min (for review, see Iwamoto et al., 2015). This rapid turnover strongly suggests that instead of influencing long-term epigenetic memory, acetylation participates in short-term alterations of chromatin states (Barth and Imhof, 2010). Homeostasis of histone acetylation is maintained by histone acetyl transferases (HATs), which catalyze the histone acetylation, and histone deacetylases (HDACs), which remove the acetyl group (Pandey et al., 2002).

Histone Lys methylation is another well-studied epigenetic modification with both activating and repressing roles in gene expression. Histone Lys residues can be mono-, di-, or trimethylated. Each distinct methylation state confers different functional outcomes depending on the methylated histone residue, the degree of methylation, and the chromatin context (Liu et al., 2010). Methylated lysines constitute landmarks for binding of chromo, MBT, Tudor, W40 domain, or PHD fingers containing proteins, which subsequently recruit additional protein complexes resulting in the compaction or relaxation of chromatin structure (Greer and Shi, 2012).

Compared to yeast and animals, much less is known about the repertoire of histone modification sites in plants. Pioneering work in *Arabidopsis* includes genome-wide chromatin indexing by ChIP (Roudier et al., 2011; Sequeira-Mendes et al., 2014) and characterization of histone modifications by mass spectrometry (Johnson et al., 2004; Bergmüller et al., 2007; Zhang et al., 2007). In order to expand this knowledge, we implemented HPLC histone separation in combination with mass spectrometry to identify uncharacterized modifications on the N-terminal tail of histone H3 in *Arabidopsis* and its close relative cauliflower (*Brassica oleracea*). We found H3K36 acetylation (H3K36ac) as a novel histone modification in plants. *GCN5*, a HAT, and HDA19, an HDAC, contribute to H3K36ac homeostasis. H3K36ac is enriched in euchromatic regions and excluded from heterochromatin. In addition, H3K36ac is generally found at the 5' end of transcriptionally active genes where it overlaps with H3K4me3 and the H2A.Z histone variant. As H3K36 can also be mono-, di-, or trimethylated, we tested if the mechanisms of trimethylation or acetylation of H3K36 are interdependent. H3K36ac was increased in a mutant with reduced H3K36me3, and H3K36me3 was increased in *gcn5* mutants with reduced H3K36ac. Thus, acetylation and methylation at H3K36 are in a competitive equilibrium that defines two alternative states.

RESULTS

The Histone H3 Modification Profile Is Conserved in Brassicaceae

A histone extract from cauliflower heads was subjected to reverse-phase HPLC, and fractions corresponding to

H3 based on immunoblot detection were collected and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using an LTQ-Orbitrap XL instrument. For the analysis of long peptides both collision-induced dissociation (CID) and electron transfer dissociation (ETD) fragmentation were performed (Supplemental Fig. S1). The high accuracy achieved in the MS measurements allowed to distinguish between trimethylated and acetylated peptides ($\Delta m = 0.0364$ D). Data were submitted to the Mascot search engine for identification of proteins and posttranslational modifications, with a focus on acetylation and methylation. This produced a map of H3 acetylation and methylation in cauliflower inflorescences (Supplemental Table S1). Most of the identified H3 modifications and their combinations are consistent with those in *Arabidopsis* (Johnson et al., 2004; Zhang et al., 2007), suggesting a high conservation of the histone modifications.

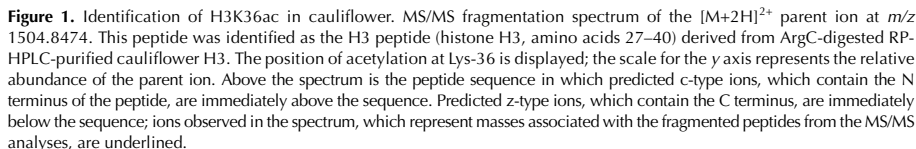
H3K36ac Is a New Modification in Arabidopsis

We identified H3K36ac as a novel H3 modification in plants (Supplemental Table S1). H3K36ac was found in two independent experiments using different enzymes (Chymotrypsin and ArgC). The confidence for the localization of the modification site was higher than 90%. The tandem mass spectrometry (MS/MS) spectrum of the ArgC-digested histone H3 shown in Figure 1A details the c and z fragmentation ions for the parent peptide with an observed mass of m/z 1504.8468. The accurate mass of the recorded parent ion was consistent with peptide acetylation ($\Delta M = 42.0106$ D) and not trimethylation ($\Delta M = 42.0470$ D). Importantly, we detected H3K36 mono-, di-, and trimethylation on other H3 peptides, confirming that this amino acid can be either methylated or acetylated (Fig. 1A; Supplemental Figs. S2 and S3).

To validate and extend the mass spectrometry results, a histone extract from *Arabidopsis* inflorescences was separated by PAGE and probed with an anti-H3K36ac antibody. The antibody was previously tested for specificity in dot blots and protein immunoblots using modified and unmodified peptides and validated for ChIP-seq (Morris et al., 2007; Egelhofer et al., 2011). The antibody showed no cross-reactivity with 11 different tested acetylated histone peptides (H2AK5ac, H3K14ac, H3K18ac, H3K27ac, H3K4ac, H3K9ac, H4K12ac, H4K16ac, H4K5/8/12/16ac, H4K5ac, and H4K8ac) and many unmodified or methylated histone peptides. The positive signals for *Arabidopsis* histone extracts (Fig. 2A) confirmed the mass spectrometric identification of the H3K36ac modification in plants. Together, proteomic and molecular methods support the identification of H3K36ac as a novel histone modification in *Brassicaceae*.

H3K36ac Is Present in Euchromatin

To establish the distribution of H3K36ac in the nucleus, paraformaldehyde-fixed *Arabidopsis* root tip nuclei were immunostained using the anti-H3K36ac



the more compact chromatin structure in heterochromatin (Rehrauer et al., 2010; Shu et al., 2012). Together, these results suggest that H3K36ac is present on active genes in euchromatin but not on heterochromatic loci.

We performed a native ChIP-seq experiment to determine the genome-wide localization of H3K36ac at mononucleosome resolution (Supplemental Table S2). To control for nucleosome density, H3K36ac reads were normalized to the anti-H3 read counts. After mapping to the five Arabidopsis chromosomes, the highest

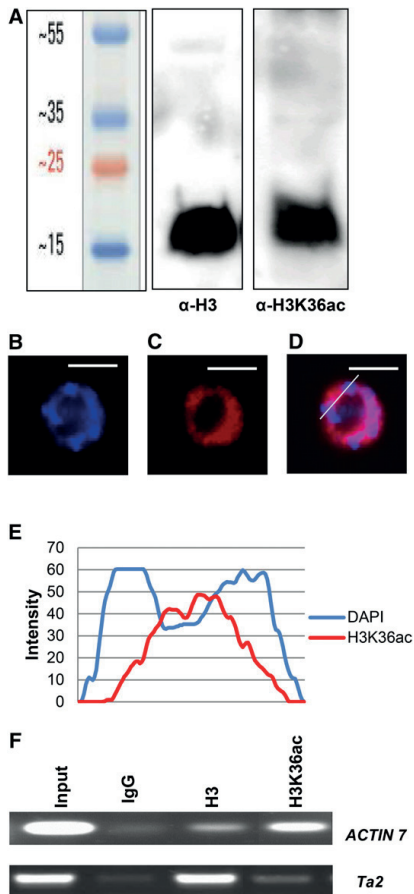


Figure 2. Detection of H3K36ac in Arabidopsis. **A**, Acid-extracted histones from Arabidopsis inflorescences were resolved on a 15% SDS-PAGE gel, followed by protein immunoblotting using anti-H3 and anti-H3K36ac antibodies. **B** and **C**, Localization of H3K36ac in root tip nuclei. Root tip interphase nuclei of Arabidopsis plants were stained with DAPI (**B**) and analyzed for immunofluorescence using an anti-H3K36ac antibody (**C**). **D**, The overlay of DAPI and H3K36ac signal. The line indicates a section through the nucleus that was analyzed using ImageJ software. **E**, Quantitative line profiles of DAPI (blue) and H3K36ac (red) fluorescence intensities. Bars = 5 μ m. **F**, X-ChIP assays were performed using IgG, anti-H3, and anti-H3K36ac antibodies with primers against the active *ACTIN7* gene and the silent *Ta2* transposable element.

enrichment was found along chromosome arms. In contrast, pericentromeric heterochromatic regions were considerably less enriched (Supplemental Fig. S4). Furthermore, the H3K36ac coverage profile was similar to the gene density distribution, consistent with an enrichment of H3K36ac at or near genes. These results support the H3K36ac immunolocalization data. To

explore the distribution of H3K36ac along gene bodies in more detail, averaged coverage profiles were plotted around transcriptional start sites (TSSs) and transcriptional termination sites (TTSs) for all annotated gene models (Fig. 3A). The results show a strong enrichment of H3K36ac between 35 and 516 bp downstream of the TSS (positions with 50% of maximal peak height). In contrast, H3K36ac was not enriched in promoter regions, in gene bodies beyond the first 500 bp distal to the TSS, or around TTSs.

Next, we tested whether the width of the H3K36ac peak depended on gene length by plotting averaged H3K36ac profiles separately for bins of increasing gene length (Fig. 3B). This analysis revealed that position, shape, and width of the H3K36ac enrichment peak are independent of gene length and that H3K36ac covered only about 480 bp downstream of the TSS. This invariant pattern of the H3K36ac modification allowed the calculation of a gene-specific H3K36ac score as the mean H3K36ac signal within a TSS-distal 35- to 516-bp window. The H3K36ac score has a bimodal distribution (Fig. 3C) that allowed to separate 16,431 genes with low H3K36ac (score ≤ 27) from 10,920 genes with high H3K36ac (score > 27). To avoid confounding effects of alternative transcripts, only genes with a single annotated transcript were included in this analysis. Since H3K36ac is found at genes and the differential enrichment of this modification allows the clear distinction of two gene classes, we asked if its presence is a binary signal for transcription. Globally, genes with higher H3K36ac scores had also higher mRNA levels (Fig. 3D). Only active but not inactive genes had the pronounced H3K36ac peak near the TSS (Fig. 3E). However, the relationship between H3K36ac and transcript levels is not linear but strongly sigmoidal (red trend line in Supplemental Fig. S5), indicating that H3K36ac levels are only weakly proportional to transcription rate and instead are a qualitative indicator of transcriptional activity. This interpretation is consistent with the bimodal distribution of H3K36ac scores (Fig. 3C).

Since H3K36ac is associated with gene activity, we analyzed its relationship to other chromatin marks associated with transcription, namely, H3K4me3, H3K36me2, H3K36me3, and the histone variant H2A.Z. Similar to H3K36ac, position, shape, and width of the H3K4me3 and H2A.Z enrichment peaks are independent of gene length (Supplemental Fig. S6). In contrast, H3K36me3 and H3K36me2 profiles depend strongly on gene length. Next, averaged H3K4me3, H3K36me2, H3K36me3, and H2A.Z profiles were plotted separately for five bins of increasing H3K36ac scores. All tested chromatin marks showed a positive correlation with H3K36ac, but the correlation with H3K4me3 and H3K36me3 was most pronounced (Supplemental Fig. S7). H3K4me3, H3K36me3, and H2A.Z are restricted to the 5'-end of genes, while H3K36me2 increases toward the 3'-end. To better compare peak positions, averaged profiles for the analyzed marks were plotted together (Fig. 4). This analysis revealed that H3K36ac and H2A.Z are localized closest to the TSS and likely associated

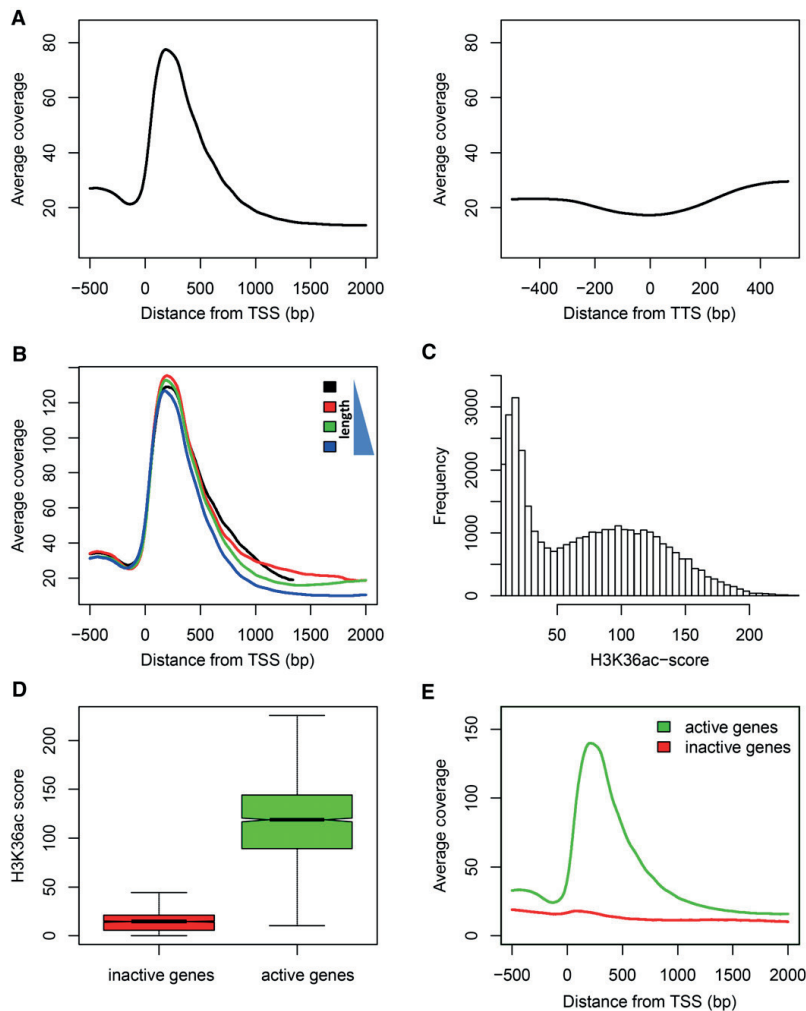


Figure 3. H3K36ac distribution across genes. A, Metagenetic plots of H3K36ac enrichment around the TSS (left panel) and TTS (right panel). The highest enrichment was detected within the first 500 bp distal to the TSS. B, H3K36ac enrichment is independent of gene length. Metagenetic plots are shown for the first (black), second (red), third (green), and fourth (blue) length quartile where the black line represents the 25% shortest genes. C, Distribution of gene-specific H3K36ac scores defined as the mean H3K36ac signal within a window of 35 to 516 bp distal to the TSS. Note the two separate populations of genes with scores of less or more than 27, respectively. D, Relation between H3K36ac score and transcriptional state. H3K36ac scores are shown for the 10% genes with lowest (red) and highest (green) mRNA levels, respectively. E, H3K36ac metagenetic plot around the TSS for the 10% genes with lowest (red) and highest (green) mRNA levels, respectively. Only active genes have the characteristic H3K36ac peak within the first 500 bp. Expression data are from Shu et al. (2012).

with transcription initiation and the transition to transcription elongation (H3K36ac peaks at 199 bp and H2A.Z at 252 bp distal to the TSS). Downstream of the H2A.Z peak, H3K4me3 and H3K36me3 (peaks at 368 and 629 bp, respectively) become more dominating

until H3K36me2, which is associated with elongating RNA Polymerase II (Oh et al., 2008), increases toward the 3'-end of genes.

Profiling of H3K36ac specifically along active genes showed the highest enrichment around the TSS irrespective

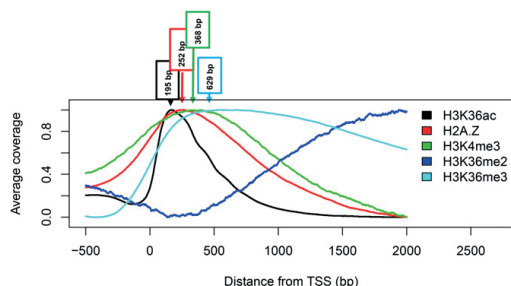


Figure 4. Metagene plots of active chromatin marks around the TSS. Phased enrichment of H3 modifications and the histone variant H2A.Z around the TSS. The peak of enrichment is indicated for each profile. Data are from Reed et al. (1996), Zilberman et al. (2008), and this study.

of feature length (Fig. 5, A and B), consistent with the distribution obtained when all genes were used (Fig. 3A). Enrichment of H3K36me3, on the other hand, also reaches a peak right after the TSS but then extends further toward the 3'-end (Fig. 5, C and D). Interestingly, we observed that the enrichment of H3K36me3 along the whole gene body increased with gene length (Fig. 5D). This is consistent with the model that H3K36me3 is associated with Polymerase II processivity (Oh et al., 2008).

Together, H3K36ac is most enriched on +1 and +2 nucleosomes of the TSS of transcriptionally active genes and correlates with H3K4me3, H3K36me2, H3K36me3, and the histone variant H2A.Z.

H3K36ac Depends on GCN5 in Arabidopsis

Histone acetylation is established by HATs. Arabidopsis has 12 different HAT genes that are grouped into four classes (Pandey et al., 2002), of which *GCN5* (GENERAL CONTROL NON DEPRESSIBLE PROTEIN5) is highly expressed in most sporophytic tissues (Zimmermann et al., 2004). In yeast, *GCN5* acetylates H3K14 (Roth et al., 2001). Because *GCN5* is conserved in plants and because the H3 sequences around K14 and K36 are similar (STGGK¹⁴AP versus STGGVK³⁶KP), we tested if Arabidopsis *GCN5* acetylates H3K36 using the *gcn5-1* mutant for X-ChIP assays. The TSS regions of two genes with high H3K36ac levels were analyzed using quantitative PCR. Their H3K36ac levels were reduced more than 5-fold in *gcn5-1* (Fig. 6A), indicating that *GCN5* is required for full H3K36 acetylation in vivo but that other HATs can also acetylate H3K36 explaining the residual H3K36ac in *gcn5-1*.

To test whether *GCN5* can directly acetylate H3K36, in vitro HAT assays were performed using N-terminal H3 fragments and purified recombinant GST and GST-*GCN5*. Acetylated histone fragments were probed with anti-H3K36ac antibody. A strong signal was obtained when using *GCN5* enzyme, while no signal was obtained with control GST or H3 alone (Fig. 6B). The in

vivo and in vitro assays therefore establish Arabidopsis *GCN5* as an H3K36 acetyl transferase and strongly suggest that it is the main enzyme depositing H3K36ac in vivo.

Competing Modifications Targeting H3K36

H3K36 can also be mono-, di-, or trimethylated at its ϵ -amino group. The Arabidopsis histone Lys methyltransferase SET DOMAIN GROUP8 (SDG8) deposits di- and trimethylation on H3K36 (Zhao et al., 2005; Xu et al., 2008). Considering that *GNC5* and *SDG8* both modify H3K36, we tested the mechanistic dependency of acetylation and methylation at H3K36. We performed X-ChIP-seq using the anti-H3K36ac antibody in *sdg8-2* and an anti-H3K36me3 antibody in *gcn5-1* because in our experience X-ChIP is more robust than N-ChIP when comparing genotypes. Crosstalk between *GCN5* and *SDG8* could result in reciprocal effects on H3K36ac and H3K36me3 enrichments, respectively. To avoid confounding effects from inactive genes, we restricted our

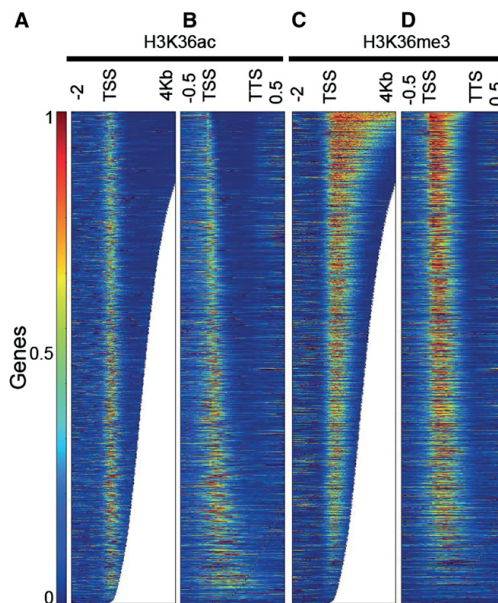
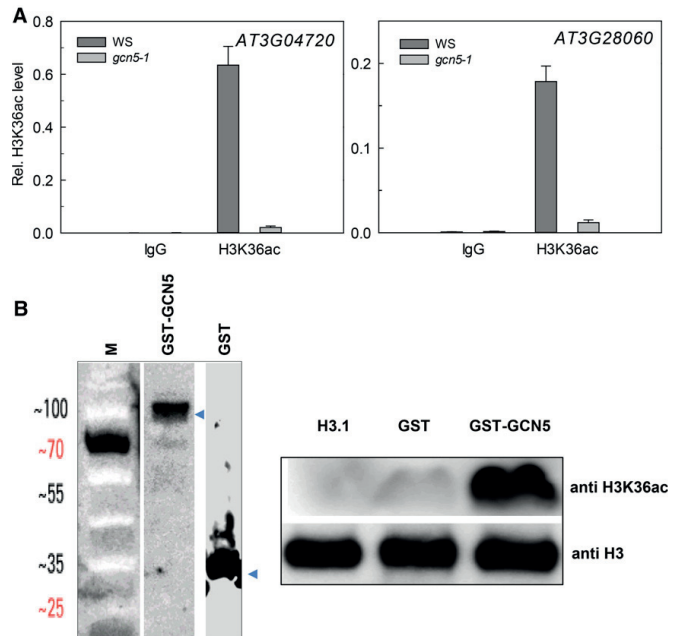


Figure 5. H3K36ac and H3K36me3 distribution across transcribed genes. A and B, H3K36ac enrichment along transcribed genes is shown on an absolute scale from 2 kb upstream to 4 kb downstream of the TSS (A) and on a relative scale from TSS to TTS (B). H3K36ac was mainly found in the first 500 bp of the gene body. The enrichments are similar in level and breadth irrespective of the length and the fraction of the gene being covered. C and D, H3K36me3 enrichment along transcribed genes is shown on an absolute scale from 2 kb upstream to 4 kb downstream of the TSS (C) and on a relative scale from TSS to TTS (D). H3K36me3 localizes close to the TSS and expands further toward the 3'-end. Longer genes have higher and wider levels of H3K36me3. Genes were ordered by length.

Figure 6. GCN5 is required for H3K36 acetylation. A, H3K36ac levels on *AT3G04720* and *AT3G28060* depend on GCN5. Ten-day-old seedlings were used for X-ChIP with anti-H3 and anti-H3K36ac antibodies. The enrichment of H3K36ac relative to H3 is shown as mean \pm se. B, H3K36ac acetyltransferase activity of GCN5. Left: Recombinant GST and GST-GCN5 were purified with Glutathione Sepharose 4B resin, blotted, and probed with anti-GST antibodies; M, weight marker. Right: HAT assays were performed using recombinant His-H3 as substrate. The reaction products were resolved on a 15% SDS-PAGE gel, transferred to polyvinylidene difluoride membrane, and probed with anti-H3K36ac antibodies. Anti-H3 was used as control.



analysis to active protein-coding genes (see Methods). Globally, H3K36ac levels in genes increased in *sdg8-2*, in particular close to the TSS (Fig. 7A). In addition, H3K36ac increased close to the TTS of genes in which its levels are normally low. Similarly, enrichment of H3K36me3 increased in *gcn5-1* and was highest in the center of genes (Fig. 7B). This increase in enrichment of acetylation and methylation in *sdg8-2* and *gcn5-1*, respectively, indicates that both modifications affect each other. Because the differences were found in distinct regions along gene bodies, this crosstalk probably does not reflect a direct competition for the H3K36 but an indirect functional repressive interaction between the acetylation and methylation enzymes.

HDA19 but Not HDA6 Affects H3K36ac Levels

Eighteen different HDACs have been identified in Arabidopsis and grouped into three different classes (Pandey et al., 2002). To identify HDACs responsible for H3K36 deacetylation, we tested the possible involvement of HDA6 and HDA19, which are two major Arabidopsis HDACs. HDA6 removes acetyl groups from H3K9ac, H3K14ac, H3K18ac, H3K23ac, H3K27ac, H4K5ac, H4K8ac, and H4K12ac (for review, see Hollender and Liu, 2008). To test if HDA6 deacetylates H3K36ac, we performed X-ChIP assays with wild-type, *hda6-1*, and *35S::HDA6-FLAG* plants using two genes with high H3K36ac levels in the wild type. No significant differences of H3K36ac levels were found in HDA6

loss-of-function or overexpressing plants (Supplemental Fig. S8), suggesting that HDA6 does not deacetylate H3K36 in vivo at least at these two loci. In contrast, a considerable increase of H3K36ac levels was found for the same genes in an *HDA19* RNAi line (Fig. 8), indicating that HDA19 is one of the Arabidopsis HDACs that deacetylate H3K36ac.

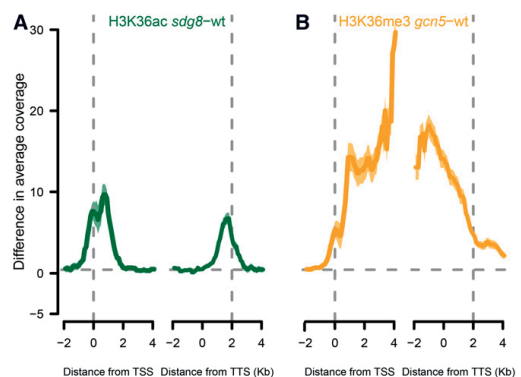


Figure 7. Competing modifications of H3K36 by GCN5 and SGD8. A, Gain of H3K36ac along genes in the *sdg8-2* background. B, Gain of H3K36me3 along genes in the *gcn5-1* background. Shaded areas represent 95% confidence intervals.

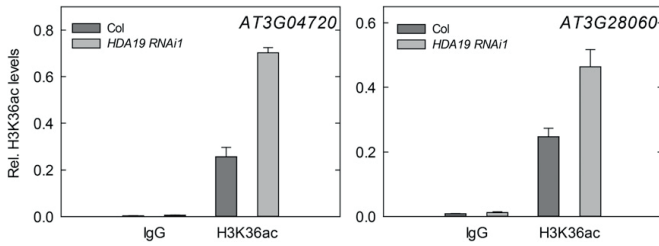


Figure 8. HDA19 affects H3K36ac levels. H3K36ac levels on *AT3G04720* and *AT3G28060* are affected by HDA19 down-regulation. Ten-day-old seedlings were used for X-ChIP with anti-H3 and anti-H3K36ac antibodies. The enrichment of H3K36ac relative to H3 is shown as mean \pm se.

H3K36ac Is a Conserved Histone Modification in Plants

Histones are among the most highly conserved eukaryotic proteins, and K36 is found in H3 of mouse, yeast, and plants such as *Arabidopsis*, tobacco (*Nicotiana tabacum*), wheat (*Triticum aestivum*), rice (*Oryza sativa*), and Norway spruce (*Picea abies*; Fig. 9A). To test if H3K36ac exists only in *Brassicaceae* or also in other plant species, histones were extracted from tobacco, rice, wheat, and spruce as representatives of monocotyledonous and dicotyledonous angiosperms as well as gymnosperms. The presence of H3K36ac was analyzed using protein immunoblotting with anti-H3K36ac antibody. A distinct H3K36ac band was visible in all histone preparations (Fig. 9B), indicating that H3K36ac is a highly conserved modification in seed plants and likely has a conserved biological function.

To test whether the H3K36ac enrichment on active genes is conserved in other plant species, X-ChIP was performed using Norway spruce cell culture. Two loci with low expression (*MA_16054g0010* and the *EnSpm* transposon-like locus *MA_10270349*) and two loci with high expression (*MA_491379g0010* and *MA_94077g0010*) were selected for ChIP analysis (Supplemental Fig. S9) using two primer sets for each gene: one set probing a region upstream of the transcriptional unit and the other set probing a region downstream of the TSS in the gene body near the start codon. Similar to the results in *Arabidopsis*, strong H3K36ac ChIP signals were present at the active genes *MA_491379g0010* and *MA_94077g0010*, but no or only weak signals were present at the silent or lowly expressed genes (Fig. 9C). In addition, H3K36ac was much higher enriched distal of the TSS than in the promoter region. Together, these results suggest that H3K36ac is present in the 5'-region of active genes in the angiosperm *Arabidopsis* and the gymnosperm Norway spruce revealing a wide conservation of this modification among seed plants.

DISCUSSION

Applying the sensitive and specific CID/ETD LC-MS/MS method (Syka et al., 2004) to histone extracts from cauliflower inflorescences, we identified a novel H3 acetylation on Lys-36 in plants. CID/ETD LC-MS/MS allows sequencing of long, highly charged peptides (>20 amino acids) and even of intact proteins and has

been successfully used already to determine histone PTMs (Taverna et al., 2007; Xiong et al., 2010; Jufvas et al., 2011). Our analysis confirmed most of the known *Arabidopsis* Lys acetylation and methylation modifications. We did not detect phosphorylated H3 peptides, which require phospho-peptide enrichment during histone extraction (Zhang et al., 2007). However, our approach identified previously unreported plant histone modifications, including H3K36ac that we analyzed in more detail. A complete list of the identified H3 post-translational modification is reported in Supplemental Table S1.

H3K36ac was previously found in yeast, *Tetrahymena*, and mammals (Morris et al., 2007). Using a high specificity

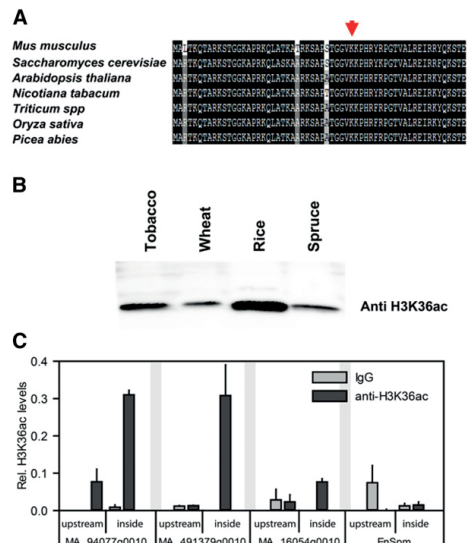


Figure 9. H3K36ac is conserved in seed plants. A, Amino acid alignment of the N-terminal tails of H3 from mouse, yeast, and several plants shows conservation of Lys-36. B, Acid-extracted histones prepared from tobacco, wheat, rice, and Norway spruce were resolved on 15% SDS-PAGE gels followed by immunoblotting using anti-H3K36ac antibodies. C, H3K36ac X-ChIP in Norway spruce. Shown are mean \pm se ($n = 3$). H3K36ac is found downstream (downstr.) of the TSS on transcribed genes.

antibody, we confirmed the MS/MS identification of H3K36ac in plants. In addition to cauliflower, H3K36ac was also found in several other seed plants. The presence of H3K36ac in yeast, ciliates, animals, and plants suggests an evolutionary conserved function.

In Arabidopsis, H3K36ac is associated with euchromatin and excluded from heterochromatic chromocenters. It is enriched within the first 500 bp and particularly at +1 and +2 nucleosomes distal to the TSS of transcriptionally active genes, irrespective of their lengths. This presence of H3K36ac distal to the TSS of Arabidopsis genes is in contrast to yeast, in which H3K36ac is highest at promoters but mostly absent from transcribed regions (Morris et al., 2007). The restriction of H3K36ac to the 5'-end of genes and the positive correlation with gene expression suggests that it is associated with the early phase of transcription but not with elongation. Although H3K36ac marks active genes, its levels are not significantly correlated with transcription rates. H3K36ac rather marks transcribed genes in a qualitative manner with full acetylation reached even at moderate transcription rates.

The peak of H3K36ac enrichment in Arabidopsis coincided well with the H2A.Z peak at the 5'-end of gene bodies. It is not clear whether the different width of the H2A.Z and H3K36ac peaks is genuine or the result of the different ChIP methods in our study and the report for H2A.Z (Zilberman et al., 2008). The coincidence of H2A.Z and H3K36ac suggests that nucleosomes in which H3K36 is acetylated often contain the histone variant H2A.Z. It remains to be established, however, whether incorporation or presence of H2A.Z favors acetylation of H3K36 and/or H3K36ac favors H2A.Z incorporation. Similar to H3K36ac and H2A.Z, H3K4me3 and H3K36me3 peak at 5'-ends of transcriptionally active genes. In the case of H3K4me3, the peak width is independent of gene length but the peak summit is shifted to the TTS. The H3K36me3 peak is shifted even further toward the 3'-end than the H3K4me3 peak. The extension of the H3K36me3 peak strongly depends on gene length. H3K36me2 increases downstream of the H3K36me3 peak and reaches its maximum only at the 3'-end. The gradient of H3K36ac, H2A.Z, H3K4me3, H3K36me2, and H3K36me3 along transcribed genes provides a spatial map of histone modifications along transcription units.

Histone acetylation is regulated by antagonistic HAT and HDAC activities. The Arabidopsis HAT GCN5 acetylates H3K36 in vitro and is necessary to achieve full H3K36ac levels in vivo. These results are consistent with the reported activity of GCN5 in yeast (Morris et al., 2007), supporting a conserved mechanism for H3K36 acetylation in eukaryotes, although as described above, the H3K36ac patterns differ between yeast and plants (Morris et al., 2007).

The Arabidopsis genome encodes 18 HDAC proteins that are grouped into three main subfamilies (RPD3, HD2, and SIR2 family) (Pandey et al., 2002). We found that H3K36ac was increased in *HDA19* RNAi lines, suggesting that *HDA19* is required for H3K36ac

homeostasis on target genes in vivo. It remains to be established whether other HDACs also modulate H3K36ac levels.

Despite the evolutionary conservation of H3K36 modifications, animals and yeast seem to use them in functionally different mechanisms than plants. In animals and yeast, H3K36me3 is associated with transcriptional elongation peaking toward the 3'-end, and H3K36ac in yeast is restricted to promoters. In plants, H3K36me3 peaks in the 5'-half of the gene, and H3K36ac is restricted to an approximately 480 bp distal to the TSS. In Arabidopsis, H3K36ac and H3K36me3 are found in overlapping regions downstream of the TSS. Our genome-wide analysis revealed a negative crosstalk between the HAT GCN5 and the HKMT SDG8. This resulted in increased H3K36ac levels when H3K36me3 is reduced and vice versa; however, the distinct spatial patterns of H3K36ac and H3K36me3 were not affected. Further work will be required to understand the mechanistic basis for this crosstalk. It is tempting to speculate that acetylation and methylation at H3K36 form a molecular switch, in which one of the two dominates. However, since nucleosomes have two H3 molecules, concomitant acetylation and methylation at H3K36 in the same nucleosome are possible. It remains to be shown whether such nucleosomes with asymmetrical K36 modifications do indeed exist.

MATERIALS AND METHODS

Plant Material

The following Arabidopsis (*Arabidopsis thaliana*) lines were previously described: *HDA19-RNAi1* (Zhou et al., 2005), *gcn5-1* (Vlachonasis et al., 2003), *35S::Flag-HDA6* (Earley et al., 2006), *hda6-1* (Aufsatz et al., 2002), and *sdg8-2* (Xu et al., 2008). The *HDA19-RNAi1* was shown to have no off-target effects on the close *HDA19* homolog *HDA6* (Zhou et al., 2005; Mehdi et al., 2016). Seeds were sown on 0.5× basal salts Murashige and Skoog medium (Duchefa), stratified at 4°C for 1 d, and then transferred to growth chambers for germination at 20°C under long-day (16 h light) photoperiods for 10 d. The seedlings were transferred to pots containing soil and cultivated in growth chambers under long-day conditions. Tobacco (*Nicotiana tabacum*), wheat (*Triticum aestivum*), and rice (*Oryza sativa*) were grown in a greenhouse at 22°C in long days. Cauliflower (*Brassica oleracea*) was obtained from a local supermarket; Norway spruce (*Picea abies*) needles were collected from an approximately 15-year-old wild-grown tree in a forest close to Uppsala (Sweden) located at 59.81° N, 17.66° E. For ChIP and quantitative RT-PCR in Norway spruce, embryogenic calli (line 11:18:1:1) derived from a zygotic embryo were used (von Arnold and Clapham, 2008).

Histone Extraction and Purification from Plants

To extract histones from inflorescences of cauliflower and Arabidopsis or from fully developed leaves of tobacco, wheat, rice, and spruce, approximately 3 g of material were ground to a fine powder and homogenized for 15 min in histone extraction buffer (0.25 M Suc, 1 mM CaCl₂, 15 mM NaCl, 60 mM KCl, 5 mM MgCl₂, 15 mM PIPES, pH 7.0, 0.5% Triton X-100 including protease inhibitors cocktail [Roche], and 10 mM sodium butyrate). After centrifugation for 20 min at 4°C and 4,500g, pellets were dissolved in 0.1 M H₂SO₄. After further centrifugation for 10 min at 17,000g, total histones were precipitated from the supernatant with concentrated trichloroacetic acid to a final concentration of 33%. The histone pellet was washed twice with acetone-0.1% HCl, air dried, dissolved either in 1× Laemmli buffer and subjected to SDS-PAGE, followed by protein immunoblotting using anti-H3 (07-690; Upstate/Millipore) and anti-H3K36ac (39379; ActiveMotif) antibodies, or fractionated by reverse-phase HPLC.

Histone Fractionation by Reverse-Phase HPLC and PTM Identification by MS/MS

Bulk histones from cauliflower were prepared as described above. The pellet was dissolved in 100 μ L of water, sonicated for 1 min, and centrifuged for 1 min at 10,000g. The supernatant was gently aspirated by pipetting and loaded on an ECLIPS XDB-C8 (4.6 \times 150 mm; Agilent) connected to an Agilent HP1100 binary HPLC system. Histone variants were separated and eluted according to Tweedie-Cullen et al. (2009). The fractions containing H3 were combined, dried, and dissolved in 50 mM ammonium bicarbonate, pH 8.0. After reduction with DTT (10 mM) and alkylation with iodoacetamide (40 mM), histones H3 were digested for 2 h at 37°C with Chymotrypsin (enzyme:substrate ratio 1:50 [w/w]; Promega) or with Arg_C (1:50 [w/w]; Roche). Samples were desalted with C18 ZipTips (Millipore) following the manufacturer's recommendations and dried. Peptides were resolubilized in 10 μ L of 5% acetonitrile and 0.1% trifluoroacetic acid and were analyzed by LC-MS/MS on an LTQ-Orbitrap XL-ETD mass spectrometer (Thermo Fischer Scientific) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies). Solvent composition at the two channels was 0.2% formic acid and 1% acetonitrile for channel A, and 0.2% formic acid and 80% acetonitrile for channel B. Peptides were resolubilized in 10 μ L of 5% acetonitrile and 0.1% formic acid and loaded on a self-made tip column (75 μ m \times 70 mm) packed with reverse-phase C18 material (AQ, 3 μ m 200 Å; Bischoff). Elution was performed with a flow rate of 200 nL/min by a gradient from 3 to 10% of B in 6 min, 35% B in 38 min, and 97% B in 45 min. One scan cycle comprised of a full-scan MS survey spectrum followed by up to six sequential CID and ETD MS/MS scans on the three most intense signals above a threshold of 500. Full-scan MS spectra (300–2000 m/z) were acquired in the FT-Orbitrap with a resolution of 60,000 at 400 m/z after accumulation to a target value of 200,000, while CID and ETD MS/MS spectra were recorded in the linear ion trap (target value of 1e4 for ion trap MSn scans). The ETD anion target value was set at 3e5 and activation time at 50 ms. Supplementary activation was employed to enhance the fragmentation efficiency for 2+ precursors and charge state-dependent ETD time enabled. The ETD reaction time was 120 ms and isolation width was 3 m/z . Charge state screening was enabled and singly charged states were rejected. Precursor masses selected twice for MS/MS were excluded for further selection for 120 s. Samples were acquired using internal lock mass calibration set on m/z 429.0887 and 445.1200.

Mass spectrometry raw files were converted into Mascot generic files (mgf) with Mascot Distiller software 2.4.2.0 (Matrix Science). Peak lists were searched using Mascot Server 2.3 against the Arabidopsis TAIR10 protein database. Search parameters were as follows: requirement for chymotryptic or ArgC ends, one missed cleavage allowed, peptide tolerance \pm 6 ppm, and MS/MS tolerance \pm 0.5 D. Carbamidomethylation of Cys was set as fixed modification, while oxidation (Met), acetylation (N-term protein, Lys), and mono-, di-, and trimethylation (Arg and Lys) were set as variable. To assess the location of the posttranslational modification sites, we used the site localization analysis provided by Mascot. The presence of modification sites was further validated by manual inspection of spectra.

ChIP and ChIP-Seq

Native ChIP and X-ChIP assays were performed as described (Bernatavichute et al., 2008; Exner et al., 2009) using gene-specific primers (Supplemental Table S3). See supplemental information for details. For ChIP-seq, leaf number six of plants grown for 35 d under short-day conditions (8 h light) was harvested at Zeitgeber time = 7 and used for nuclei isolation. Supplemental Table S2 gives a summary of the ChIP-seq experiments.

Bioinformatics and ChIP-Seq Data Analysis

Illumina reads were mapped to the Arabidopsis reference TAIR10 genome using bowtie2 (version 2.1; Langmead and Salzberg, 2012). SAM file output from bowtie was converted to BAM format using SAMtools (version 1.4; Li et al., 2009) and imported into R (version 2.15.2; <http://www.R-project.org/>) using functions from the Rsamtools package. All subsequent analysis was performed in R. Identical reads present more than 25 times were considered as PCR artifacts and filtered out using the filterDuplReads function from package HtSeqTools. Replicates were pooled, and sample normalization to identical sequencing depth and H3K36ac signals was normalized to histone signals to control for variable nucleosome density using functions from package nucleR (Flores and Orozco, 2011). H3K4me3, H3K36me2, H3K36me3, and H2A.Z data were taken from the literature (Oh et al., 2008; Zilberman et al., 2008). Gene

expression data are from Shu et al. (2012) and match cultivation conditions, plant age, harvest time, and tissue used here for ChIP-seq. The expression cutoff to consider a gene as active was a value of 6 in the microarray data from Shu et al. (2012), resulting in a set of 12,102 active protein coding genes. Heat maps were generated using the open source tool deepTools (Ramírez et al., 2014).

Constructs and Recombinant Protein Expression and Purification

The coding region for the N-terminal tail of H3 (amino acids 1–50) was amplified by PCR from Arabidopsis cDNA using primers LH1625 (CTATTCTCTAAAGCAACAGTTC, forward) and LH1626 (CACCATGGCTCGTACCAA, reverse). The PCR product was cloned into vector pET100-D-TOPO according to the manufacturer's recommendations (Invitrogen). pGEX4T1 was obtained from Amersham Life Science; pQE30-GCN5 was previously described (Benhamed et al., 2006). Recombinant N-terminal 6xHis tagged H3^{1–50} was expressed in *Escherichia coli* BL21-pLys Rosetta cells and purified on Talon metal affinity resin (Clontech) according to the manufacturer's recommendations. Recombinant GST and GST-GCN5 were expressed in BL21-pLys Rosetta and purified with Glutathione Sepharose 4B resin according to the manufacturer's recommendations (Amersham Life Science). Purified proteins were stored at –20°C until use in HAT activity assays.

HAT Activity Assays

HAT assays using recombinant GST-GCN5 expressed in *E. coli* were performed as follows: 1 μ g of purified His-tagged histone H3^{1–50} was mixed with purified recombinant GST-GCN5 or with GST alone in HAT buffer (0.25 μ g/ μ L acetyl-CoA [Sigma-Aldrich], 50 mM Tris-HCl, pH 8.0, 10% [v/v] glycerol, 1 mM DTT, complete EDTA-free protease inhibitor [Roche], and 10 mM sodium butyrate). Acetylation reactions were incubated for 60 min at 30°C with gentle shaking and terminated by addition of 2 \times SDS Laemmli sample buffer followed by heating for 5 min to 100°C. The reaction products were electrophoretically resolved on 15% SDS-PAGE gels and subjected to protein immunoblotting using anti-H3K36ac (Active Motif) and anti-H3 (Upstate/Millipore) antibodies.

RNA Isolation and Quantitative RT-PCR

RNA extraction and reverse transcription were performed as described previously (Alexandre et al., 2009) using gene-specific primers (Supplemental Table S3). See supplemental information for details.

Immunostaining

Immunostaining was performed as described previously (Jasencakova et al., 2000) with some modifications. See supplemental information for details.

Accession Numbers

Sequencing reads are deposited in Gene Expression Omnibus (GSE74841).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Histone separation by RP-HPLC.

Supplemental Figure S2. Identification of H3K36me2.

Supplemental Figure S3. Identification of H3K36me2.

Supplemental Figure S4. Genome-wide distribution of H3K36ac on the Arabidopsis genome.

Supplemental Figure S5. H3K36ac is a qualitative mark for transcriptional activity.

Supplemental Figure S6. Dependency of histone marks on gene length.

Supplemental Figure S7. H3K36ac positively correlates with H3K4me3, H3K36me3, and H2A.Z.

Supplemental Figure S8. HDA6 does not affect H3K36ac steady state levels on AT3G04720 or AT3G28060.

Supplemental Figure S9. Expression levels of Norway spruce genes used for ChIP.

Supplemental Table S1. List of histone H3 modifications that were identified in *B. oleracea*.

Supplemental Table S2. Summary of ChIP-seq experiments.

Supplemental Table S3. Lists of primers used.

Supplemental Methods.

ACKNOWLEDGMENTS

We thank Cecilia Wärdig for technical help. We thank Keqiang Wu, Craig Pikaard, and Dao-Xiu Zhou for kindly providing seeds of *HDA19-RNAi1*, *35S::Flag-HDA6*, *gcn5-1*, and the pQE30-GCN5 construct. We thank the FGCZ for Illumina sequencing.

Received November 18, 2015; accepted January 13, 2016; published January 13, 2016.

LITERATURE CITED

- Alexandre C, Möller-Steinbach Y, Schönrock N, Grussem W, Hennig L (2009) Arabidopsis MSI1 is required for negative regulation of the response to drought stress. *Mol Plant* 2: 675–687
- Aufsatz W, Mette MF, van der Winden J, Matzke M, Matzke AJ (2002) HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J* 21: 6832–6841
- Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21: 381–395
- Barth TK, Imhof A (2010) Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci* 35: 618–626
- Benhamed M, Bertrand C, Servet C, Zhou DX (2006) Arabidopsis GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. *Plant Cell* 18: 2893–2903
- Bergmüller E, Gehrig PM, Grussem W (2007) Characterization of post-translational modifications of histone H2B-variants isolated from *Arabidopsis thaliana*. *J Proteome Res* 6: 3655–3668
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS One* 3: e3156
- Berr A, Shafiq S, Shen WH (2011) Histone modifications in transcriptional activation during plant development. *Biochim Biophys Acta* 1809: 567–576
- Cannon CG (1955) The interactions and structure of the CONH-group in amides and polyamides. *Mikrochim Acta* 43: 555–588
- Chen ZJ, Tian L (2007) Roles of dynamic and reversible histone acetylation in plant development and polyploidy. *Biochim Biophys Acta* 1769: 295–307
- Earley K, Lawrence RJ, Pontes O, Reuther R, Enciso AJ, Silva M, Neves N, Gross M, Viegas W, Pikaard CS (2006) Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev* 20: 1283–1293
- Egelhofer TA, Minoda A, Klugman S, Lee K, Kolasinska-Zwiercz P, Alekseyenko AA, Cheung MS, Day DS, Gadel S, Gorchakov AA, et al (2011) An assessment of histone-modification antibody quality. *Nat Struct Mol Biol* 18: 91–93
- Exner V, Aichinger E, Shu H, Wildhaber T, Alfaro P, Cafisch A, Grussem W, Köhler C, Hennig L (2009) The chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 is essential for H3K27me3 binding and function during Arabidopsis development. *PLoS One* 4: e5335
- Flores O, Orozco M (2011) nucle: a package for non-parametric nucleosome positioning. *Bioinformatics* 27: 2149–2150
- Greer EL, Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* 13: 343–357
- Henikoff S (2005) Histone modifications: combinatorial complexity or cumulative simplicity? *Proc Natl Acad Sci USA* 102: 5308–5309
- Hollender C, Liu Z (2008) Histone deacetylase genes in Arabidopsis development. *J Integr Plant Biol* 50: 875–885
- Hong L, Schroth GP, Matthews HR, Yau P, Bradbury EM (1993) Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. *J Biol Chem* 268: 305–314
- Iwamoto M, Hiraoka Y, Haraguchi T (2015) The nuclear pore complex acts as a master switch for nuclear and cell differentiation. *Commun Integr Biol* 8: e1056950
- Jasencakova Z, Meister A, Walter J, Turner BM, Schubert I (2000) Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. *Plant Cell* 12: 2087–2100
- Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293: 1074–1080
- Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE (2004) Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res* 32: 6511–6518
- Jufvas A, Strålfors P, Vener AV (2011) Histone variants and their post-translational modifications in primary human fat cells. *PLoS One* 6: e15960
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, et al (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471: 480–485
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Sub-group (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079
- Liu C, Lu F, Cui X, Cao X (2010) Histone methylation in higher plants. *Annu Rev Plant Biol* 61: 395–420
- Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* 3: e328
- Luo M, Liu X, Singh P, Cui Y, Zimmerli L, Wu K (2012) Chromatin modifications and remodeling in plant abiotic stress responses. *Biochim Biophys Acta* 1819: 129–136
- Mehdi S, Derkacheva M, Ramström M, Kraleman L, Bergquist J, Hennig L (2016) MSI1 functions in a HDAC complex to fine-tune ABA signaling. *Plant Cell* 28: 42–54
- Morris SA, Rao B, Garcia BA, Hake SB, Diaz RL, Shabanowitz J, Hunt DF, Allis CD, Lieb JD, Strahl BD (2007) Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. *J Biol Chem* 282: 7632–7640
- Oh S, Park S, van Nocker S (2008) Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genet* 4: e1000077
- Pandey R, Müller A, Napoli CA, Selinger DA, Pikaard CS, Richards EJ, Bender J, Mount DW, Jorgensen RA (2002) Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res* 30: 5036–5055
- Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42: W187–W191
- Reed JW, Foster KR, Morgan PW, Chory J (1996) Phytochrome B affects responsiveness to gibberellins in Arabidopsis. *Plant Physiol* 112: 337–342
- Rehrauer H, Aquino C, Grussem W, Henz SR, Hilsen P, Laubinger S, Naouar N, Patrignani A, Rombauts S, Shu H, et al (2010) AGRONOMICS1: a new resource for Arabidopsis transcriptome profiling. *Plant Physiol* 152: 487–499
- Roth SY, Denu JM, Allis CD (2001) Histone acetyltransferases. *Annu Rev Biochem* 70: 81–120
- Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, et al (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J* 30: 1928–1938
- Sanchez R, Meslamani J, Zhou MM (2014) The bromodomain: from epigenome reader to druggable target. *Biochim Biophys Acta* 1839: 676–685
- Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C (2014) The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* 26: 2351–2366
- Shu H, Wildhaber T, Siretskiy A, Grussem W, Hennig L (2012) Distinct modes of DNA accessibility in plant chromatin. *Nat Commun* 3: 1281

- Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403**: 41–45
- Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* **101**: 9528–9533
- Taverna SD, Ueberheide BM, Liu Y, Tackett AJ, Díaz RL, Shabanowitz J, Chait BT, Hunt DF, Allis CD (2007) Long-distance combinatorial linkage between methylation and acetylation on histone H3 N termini. *Proc Natl Acad Sci USA* **104**: 2086–2091
- Tweedie-Cullen RY, Reck JM, Mansuy IM (2009) Comprehensive mapping of post-translational modifications on synaptic, nuclear, and histone proteins in the adult mouse brain. *J Proteome Res* **8**: 4966–4982
- Vlachonasios KE, Thomashow MF, Triezenberg SJ (2003) Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect *Arabidopsis* growth, development, and gene expression. *Plant Cell* **15**: 626–638
- von Arnold S, Clapham D (2008) Spruce embryogenesis. *Methods Mol Biol* **427**: 31–47
- Xiong L, Adhvaryu KK, Selker EU, Wang Y (2010) Mapping of lysine methylation and acetylation in core histones of *Neurospora crassa*. *Biochemistry* **49**: 5236–5243
- Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, Steinmetz A, Shen WH (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. *Mol Cell Biol* **28**: 1348–1360
- Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK (2007) Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. *PLoS One* **2**: e1210
- Zhao Z, Yu Y, Meyer D, Wu C, Shen WH (2005) Prevention of early flowering by expression of *FLOWERING LOCUS C* requires methylation of histone H3 K36. *Nat Cell Biol* **7**: 1256–1260
- Zhou C, Zhang L, Duan J, Miki B, Wu K (2005) HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in *Arabidopsis*. *Plant Cell* **17**: 1196–1204
- Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**: 125–129
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol* **136**: 2621–2632

